ELECTRON TUBE DESIGN

Electron Tube Design

Electron Tube Division
RADIO CORPORATION OF AMERICA
Harrison, New Jersey

Privately issued by
ELECTRON TUBE DIVISION
Radio Corporation of America
for use by its employees and for restricted distribution.

First Printing, 1962

Printed in U.S.A.

FOREWORD

The fifty-three articles in this book were written by RCA engineers and scientists to acquaint RCA tube engineers, particularly those of limited tube experience, with the terminology, objectives, problems, and techniques peculiar to electron tubes. The articles are intended to provide the basic principles underlying the design, manufacture, and application of the many present-day varieties of tubes.

Because this book is a compilation of articles by authors, each writing more or less independently, duplication will be encountered. In this kind of material, however, duplication is desirable. First, some duplication is necessary to help keep individual articles complete in themselves; second, some repetition may assist the reader in acquiring information faster; and, most important, different treatments of similar subjects illustrate how tube design is affected by the different design objectives and the different environments of individual authors.

In general, the treatment is non-mathematical and free of detailed descriptions which often are subject to change. For more detailed treatment, the reader is referred to the references and bibliography at the end of most articles. References to published material list the publications; those to unpublished information show a code, the key to which is available to authorized persons through RCA Electron Tube Division librarians.

Because the preparation of these articles extended over several years, some authors are no longer associated with the design activities which they discuss. In such cases, however, the authors' names are shown in the article captions with the locations where the preparation of the articles took place. Most articles involved the cooperation of associates of the authors. These contributions are gratefully acknowledged by the authors.

This book is the result of a strong management desire, supported by the Engineering Education Committees at Harrison and Lancaster, to bring the original Vacuum Tube Design book up to date. It was recognized that the earlier book had played a very important role in the orientation, training, and development of RCA tube-design engineers, that a revised edition would supplement and strengthen the programs conducted by the education committees, and that engineering supervision would benefit greatly if new engineers had access to basic tube-design know-how written by RCA authorities.

An early step in implementing the program for the revised book was the formation by Dr. G. R. Shaw, then chief engineer of the Electron Tube Division, of a steering committee under the chairmanship of J. F. Hirlinger. From the first, this committee worked closely with the education committees, who had the responsibility of choosing subject matter, selecting and appointing authors, and guiding and encouraging authors in the completion of their assignments.

Much progress was made, but a greater and more concerted effort was needed to carry the program through to prompt completion. As a result, R. S. Burnap, Manager of Commercial Engineering, on his retirement, was appointed a consultant to the RCA Electron Tube Division with his most important and essentially full-time assignment the completion of the Electron Tube Design book. With the establishment of new schedules and through the special efforts of committeemen, authors, and reviewers to complete their commitments, the job was finished.

The names of contributing committee members during the preparation of the book, together with the names of those who reviewed or edited articles follow. The contributions of these people and of the men who carried the heaviest load, the authors, are hereby acknowledged with the sincerest of thanks.

R. S. Burnap

Consultant to the Electron Tube Division

ENGINEERING EDUCATION COMMITTEES

Harrison

T. M. Cunningham	Chairman (1)
W. Zvonik	Chairman (2)
H. Waltke	Chairman (3)
J. J. Thompson	Chairman (4)
J. J. Thompson	Vice Chairman (1)
R. Rauth	Vice Chairman (2)
D. Gardam	Training Representative (1)
P. Theall	Training Representative (2)

Lancaster

L. B. Headrick	Chairman (1)
G. G. Thomas	Chairman (2)
R. E. Slough, Jr.	Training Representative (1)
J. L. Hopkins	Training Representative (2)

BOOK COMMITTEES

Harrison

T. M. Cunningham	Chairman (1)
J. J. Thompson	Chairman (2)
J. Gutowski	Vice Chairman (1)
R. J. Nelson	Vice Chairman (2)
R. Lambdin	
O. H. Schade, Jr.	
L. Scholz	

Lancaster

A. Nekut*	Chairman (1)
M. Lemeshka*	Co-Chairman
M. Petrisek*	Chairman (2)
W. N. Parker*	
A. C. Grimm	
H. A. Kaufman	
A. L. Lucarelli	
H. C. Moodey	
Y. W. Uyeda	
R. H. Zachariason	

*Member of original planning group.

TECHNICAL REVIEW COMMITTEES

Harrison

Same personnel as Book Committee

Lancaster

R. H. Zachariason	Chairman
M. V. Hoover	
W. N. Parker	
D. D. Van Ormer	

STEERING COMMITTEE

J. F. Hirlinger, Chairman	Harrison	C. A. Meyer	Harrison
R. S. Burnap	Harrison	A. G. Nekut (1)	Lancaster
R. L. Klem (1)	Harrison	W. N. Parker (2)	Lancaster
T. M. Cunningham (2)	Harrison	G. R. Shaw	Harrison
M. B. Lemeshka (1)	Lancaster	H. T. Swanson (1)	Lancaster
M. Petrisek (2)	Lancaster	C. W. Thierfelder (2)	Marion

(1) First member; (2) Succeeding member

REVIEWERS AND EDITORS

H. Baum	W. F. Griffin	M. B. Knight	A. M. Morrell	C. P. Smith
B. Becker	M. Grimes	R. W. Lambdin	G. A. Morton	J. W. Smith
M. Berg	A. C. Grimm	C. T. Lattimer	L. S. Nergaard	W. A. Smith
F. G. Block	J. Gutowski	W. F. Lawrence	R. G. Neuhauser	C. H. Thomas
B. B. Brown	C. P. Hadley	E. A. Lederer	W. N. Parker	J. J. Thompson
R. S. Burnap	A. E. Hardy	R. L. Leigh	F. Paschke	A. Trax
T. M. Cunningham	W. A. Harris	J. T. Mark	H. J. Prager	J. C. Turnbull
S. B. Deal	M. D. Harsh	E. M. McElwee	O. H. Schade, Sr.	S. Umbreit
R. W. Engstrom	L. B. Headrick	C. H. Meltzer	O. H. Schade, Jr.	D. H. Van Ormer
D. W. Epstein	T. J. Henry	P. A. Metzger	L. C. Scholz	F. S. Veith
F. G. Foerster	C. W. Horsting	C. A. Meyer	F. Scott	P. Weimer
N. S. Freedman	R. B. Janes	B. Miller	B. Sheren	M. R. Weingarten
J. Gallup	M. Jinetopulos	C. D. Mitchell	R. Shrader	J. F. Wilhelm
L. P. Garner	H. Johnson	J. Monagle	A L. Smith	R. H. Zachariason
		H. C. Moodey		

PRODUCTION COORDINATOR - B. A. Pierz

CONTENTS

Article Title	Author	Page
Emitters and Photoconductors		
Fundamentals of Electron Emission	L.S. Nergaard	1
Oxide-Coated Emitters	C.P. Hadley	28
Electron-Emission Coating for the Oxide Cathode	C.H. Meltzer & E.G. Widell	37
Nickel Base Metal for the Oxide Cathode	C.H. Meltzer & E.G. Widell	64
The Interface and Other Oxide-Cathode Impedances	E.J. Hannig	77
Tungsten, Thoriated-Tungsten, and Thoria Emitters	W.E. Harbaugh	90
Photoelectric and Secondary-Electron Emission	A.H. Sommer	99
Photoconductors for Camera Tubes	C.W. Rector	108
Phosphors		
Cathodoluminescent Materials	R.N. Summergrad	128
Application and Colorimetry of Phosphor Screens	A.E. Hardy	138
Manufacture of Cathodoluminescent Materials	J.A. Markoski	153
Space Current		
Calculation of Fields and Currents	L.C. Scholz	159
Space Current Flow	M.V. Hoover	202
General Design Considerations for Receiving Tubes		
Heaters	A.G.F. Dingwall	216
Heat Transfer in Receiving Tubes	O.H. Schade, Jr.	244
Contact Potential and Grid Currents	E.R. Schrader	266
A New Approach to the Calculation of Electron-Tube Characteristics	O.H. Schade, Sr.	278
Metals, Metallurgy, and Metal Fabrication		
Metals and Metallurgy	C.W. Horsting, et al	315
The Metallic State	C.W. Horsting	
Metals for Electron Tubes	J.J. Carrona	
Testing of Metals	C.H. Li	
Cathode Materials for Receiving Tubes	L.S. Solet	
Plate Materials (Metals)	H.G. Scheible & T.A. Sternberg	
Grids and Leads	A.J. Stoeckert	
Metal-to-Metal Joining	E.S. Thall	
Joining Metals to Glasses and Ceramics	K. Strater	
Systematic Charts of the Properties of Electron Tube Metals	S.Y. Husni	

CONTENTS (Cont'd)

Article Title	Author	Page
Metals for Vacuum Tube Construction	P.D. Strubhar	370
Principles of Metal Forming	L.S. Sloane	377
Processes Used in Metal Parts Fabrication and Finishing	H.A. Kauffman	385
Insulators, Envelope Materials, and Seals		
Glass for Receiving Tubes	J. Gallup	392
Glass and Its Properties and Seals	J.C. Turnbull & G.E. Eiwen	406
Ceramics	W.F. Lawrence	433
Ceramics and Ceramic-to-Metal Seals	M. Berg	452
Design of Mica Spacers	M. Bondy & M. Jinetopulos	465
Parts Cleaning, Coating, and Basing		
Parts Cleaning and Processing	F.I. Scott	472
Coatings, Bases, and Basing Cements	L.P. Fox	481
Exhaust and Finishing		
High-Vacuum Technology and Equipment	W.G. Henderson & A.L. Lucarelli	485
Exhaust of Receiving-Type Tubes	C.J. Pearce	506
Getters	C.H. Thomas	519
Aging and Stabilization	G. Wolfe	526
Design, Processing, and Application of Receiving-Type Tubes and Picture Tubes		
Diodes	A.P. Kauzmann	537
Small-Signal Radio-Frequency Amplifier Tubes	W.A. Harris	577
Design of Intermediate-Frequency Amplifier Tubes	M.B. Knight	599
Design of Converter and Mixer Tubes	T.J. Henry	607
Design of AM and FM Detector Tubes	T.J. Henry	622
Vacuum-Tube Oscillators	W.E. Babcock	637
Beam Power Tube Design Considerations	O.H. Schade, Jr.	656
Audio Output, Vertical-Deflection, and Horizontal-Deflection Tubes for Receiving Applications	M. Bondy	669
Black-and-White Picture Tubes	Thierfelder, et al	682
Color Picture Tube Design and Processing	H.C. Moodey & A.M. Morrell	709

CONTENTS (Cont'd)

Article Title	Author	Page
Circuit Troubles Caused by Unusual Tube Effects	W.E. Babcock	736

Design, Processing, and Application of Industrial-Type Tubes

Some Factors Affecting the Design and Application of Small Power Tubes	J.W. Gaylord	743
Large-Power-Tube Design, Processing, and Testing	W.P. Bennett	763
Gas Tube Design	H.H. Wittenberg	792
Design and Processing of Oscillograph Tubes	E.M. Smith	818
Phototube Design	J.L. Weaver	826
Camera Tube Design and Processing	B.H. Vine	838
Image-Converter Tubes	R.G. Stoudenheimer	848
Design and Processing of Storage Tubes	M.D. Harsh	862
Design and Processing of Microwave Magnetrons	F.E. Vaccaro & D.E. Nelson	876
Traveling-Wave Tube Design	E.E. Bliss	898

Fundamentals of Electron Emission

L. S. Nergaard

Princeton

ELEMENTARY CONSIDERATIONS

In the preparation of this article on electron emission, two courses seemed open: (1) to give brief descriptions of the various emission mechanisms and then to present a substantial amount of data relating to these mechanisms, or (2) to give a more detailed elementary exposition of the physical phenomena involved and then to illustrate them in some detail with a particular example, the oxide cathode. The latter alternative seemed preferable because a comprehensive literature survey can make boring reading and, more important, it was believed that an elementary discussion of the physical phenomena involved in electron emission would prepare the way for a study of the vast literature on the subject and, at the very least, provide a lexicon for the reading of this literature. Accordingly, the second course was adopted.

In the subsequent discussion, the viewpoint will be quite elementary and even mechanistic where possible. To this writer, barium oxide on a cathode is not a white coating of "gunk" prepared by a complicated bit of witchcraft. The writer has no complaint with witchcraft but he regards the end result in this case as a collection of crystallites, each containing an orderly array of very tangible spherical barium and oxygen ions stacked alternately. It is just this kind of mental imagery that the writer hopes to convey so that atoms, ions, and other elementary particles will have the reality of the balls on a pool table. Then the interactions of particles are easy to visualize and to remember. As long as the behavior conforms to classical mechanics, visualization is easy, but difficulties arise when quantum effects enter and particles display wavelike properties. These difficulties may be surmounted for the most part by considering electrons and photons as short wave trains ("wave-packets") and invoking the familiar concepts of physical optics. The quantum effects which are of concern are: (1) the dual nature of particles, (2) the quantization of energy, and (3) quantum statistics.

The Dual Nature of Particles

The wave theory of light, which accounts for the numerous interference effects observed in optics and radio propagation, was firmly entrenched when Planck showed that the spectral distribution of radiation from a black body can be accounted for by assuming that radiation occurs in energy particles (quanta) of size $h\nu$, where h is a constant (Planck's constant) and ν is the frequency of the light. When it was shown that the photoelectric threshold for metals is given by

$$h\nu_0 = \phi \tag{1}$$

where ν_0 is the lowest light frequency at which photoemission occurs and ϕ is the work function of the metal, the evidence for quanta was overwhelming. Because both the wave characteristics and particle characteristics are required to encompass optical phenomena, the particles ("photons") may be thought of as wave trains characterized by a frequency, yet localized in space so that they are in a sense particles. As noted above, the energy of a photon is

$$\epsilon = h\nu \tag{2}$$

The momentum p of a photon is equal to its energy divided by its velocity (the velocity of light c), or

$$p = \frac{h}{c} = \frac{h}{\lambda} \tag{3}$$

where λ is the wave length

This value for the momentum was precisely demonstrated by A. H. Compton in 1922.

In 1924, deBroglie suggested that electrons, which are normally considered particles, might also display wave properties. He associated a wave length

$$\lambda = \frac{h}{p} \tag{4}$$

with the electrons. This relation is the same as that found for photons. In 1925, Walter Elsasser suggested* that electrons should display diffraction phenomena. Diffraction effects in accord with the above relation were found in 1927 by Davisson and Germer and by G. P. Thomson.** Thus both electrons and photons can be visualized as wave trains.

---

*See Mott, N. F., and R. W. Gurney, Electronic Process in Ionic Crystals, Oxford University Press, London, 1950.

**The electron wave length can be written $\lambda \approx (1.1 \times 10^{-7})/V^{1/2}$ cm where V is the energy of the electron in volts. Hence, an 100-volt electron has a wave length of about 10^{-8} cm, or about a lattice spacing for a metal. It is this short wave length that gives the electron microscope its high resolution.

Quantization of Energy

The possible energy states of a system are discrete. The energy losses and gains of a system correspond to transitions between the discrete states and, thus, are also discrete. This discreteness of the energy states may be regarded as a consequence of the wave nature of particles. (See Appendix A.) In some systems, the energy states are so closely spaced that the thermal energy kT (k = Boltzmann's constant, T = absolute temperature) of the particles constituting the system can cause transitions between states. In such systems, the states may be regarded as a continuous band of states.

Quantum Statistics

In describing the equilibrium distribution of particles among possible energy states, two kinds of statistics are possible: Fermi-Dirac statistics and Bose-Einstein statistics.

Fermi-Dirac Statistics. These statistics, which apply to atoms, ions, and electrons, state that the probability that a state of energy be occupied by a particle is given by the expression

$$\frac{1}{1 + \exp\left[(\epsilon - \epsilon_F)/kT\right]} \qquad (5)$$

where ϵ is the state of energy of the electron, and ϵ_F is the Fermi-energy, the energy at which the probability of occupancy is one-half. In Appendix B it is shown that transitions in a system obeying Fermi-Dirac statistics may be regarded as bimolecular and that, except for electrons, each state may be occupied by only one particle; in the case of electrons, each state may be occupied by two electrons if their spins are oppositely directed.

Bose-Einstein Statistics. These statistics apply to photons and phonons (the energy quanta of lattice vibrations) and state that the probability that a state of energy ϵ be occupied may be expressed as

$$\frac{1}{\exp(\epsilon/kT) - 1} \qquad (6)$$

We will have no need for the Bose-Einstein statistics.

The subsequent discussion will consist of a brief introduction to solid-state physics and will lay the ground work for the discussion of electron-emission processes. The emission processes involved in thermionic emission, field emission, photo-emission, and secondary emission will then be outlined. Finally, the physics of the oxide cathode will be discussed in some detail. The oxide cathode is the bread and butter of the tube industry and particular attention to this emitter is well justified.

ELEMENTARY SOLID-STATE PHYSICS*

If all that an emitter did was to emit electrons copiously, life would be simple. However, such is not the case; an electron emitter also emits atoms and molecules as well, and may absorb atoms and molecules from ambient gases. The emission and absorption of atoms and molecules changes the constitution of the emitter and alters its emitting properties. Therefore, it is necessary to consider some properties of solids other than their electron-emitting properties if the behavior of emitters is to be understood.

Some Properties of Solids

A solid is an assemblage of atoms or ions. The parent atoms can be thought of as almost rigid spheres each consisting of a positively charged nucleus surrounded by electrons. Outside of this core of nucleus and electrons are a few electrons (the valence electrons) which determine the chemical properties of the atoms. The total charge carried by the core electrons and the valence electrons is equal to the nuclear charge so that the net charge on an atom is zero. When atoms are packed together that may share electrons, i.e., the valence electrons of one atom may spend part of the time on neighboring atoms. When this sharing occurs, the crystal is said to be covalently bound. Germanium and silicon are examples of covalent crystals. Another possibility exists when a crystal contains atoms of two or more kinds. Then, the valence electrons of one constituent may transfer to atoms of another constituent to form positive and negative ions; the ions are held together by coulomb forces. Such crystals are called ionic crystals. The alkali-halides, such as NaCl, are examples of ionic crystals. In either event, the crystal can be regarded as made up of stacked spheres. Fig. 1 shows a two-dimensional model of a crystal; each ball bearing represents an ion or an atom. This model gives an effective visualization of how a crystal is constituted and illustrates several of the defects that occur in real crystals. A lattice vacancy is obvious and the slip plane of a dislocation due to the non-uniform strain on the boundary is quite visible. Fig. 2 shows the effect of an "impurity" (an oversize atom) on the crystal. Notice that the crystal makes room for the impurity by leaving vacancies.

How well the packed-sphere model is borne out in real crystals is shown by the fact that Goldschmidt was able to select a set of numbers, one assigned to each ion, such that the sums of the numbers associated with any pair of ions gave the lattice spacing of the crystal formed from the ions.* These numbers are called ionic radii and range from 0.07 A for F^{7-} to 2.94 A for Sn^{4+}.

Although the atoms are almost rigid, they still can move somewhat and they do move with an energy of motion that depends on the temperature. The energy stored in atomic or ionic motion is responsible, in large part, for the specific heat of solids; the remainder of the specific heat is due to the heat motion of electrons.

Some effects of this heat motion are illustrated in

*The following books on solid-state physics are standard; Kittel is the latest and, in some ways, the most readable; Seitz, F., The Modern Theory of Solids, McGraw-Hill, New York, 1940; Mott, N. F., and R. W. Gurney, Electronic Process in Ionic Crystals, Oxford University Press, London, 1950; Kittel, C., Introduction to Solid State Physics, Wiley and Sons, New York, 1956.

*Tables of ionic radii are to be found in Seitz, F., The Modern Theory of Solids, McGraw-Hill, New York, 1940

Fig. 3. This is an energy diagram in which the energy of a particle is shown vertically and its position horizontally. The binding energy of the atoms is shown by the periodic curve; it shows the potential energy that must be surmounted for an atom to jump from a normal lattice site, near a minimum, to an adjacent lattice site. Such a jump can occur when: (1) an atom picks up enough thermal energy to surmount the potential hill and (2) an adjacent lattice site is vacant. This jump process is the vacancy mechanism for self-diffusion and the energy ϵ_D is the activation energy for self-diffusion. Notice that where the crystal terminates on the right, there is a potential barrier of height ΔH. When a surface atom picks up enough thermal energy to surmount this hill it is "free" and can escape from the solid. The value ΔH, then, is the heat of sublimation.

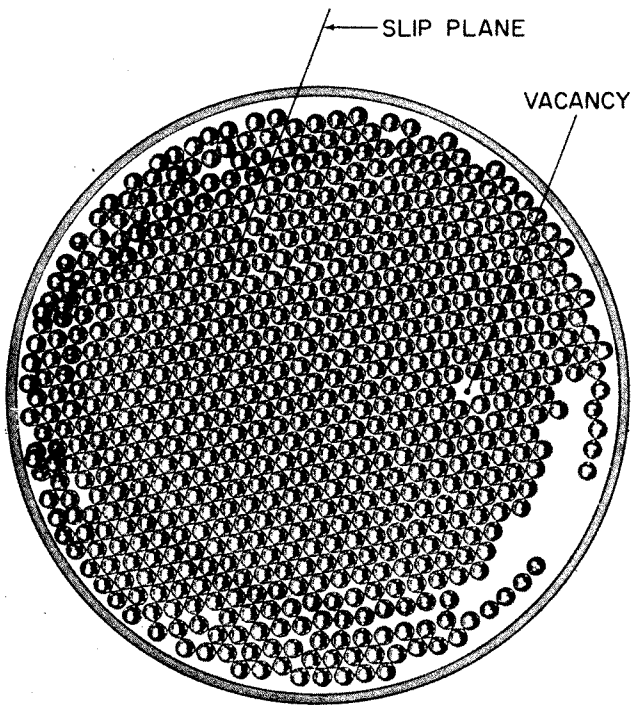


Figure 1. Ball-Bearing Model of a Two-Dimensional Crystal with Several Common Crystal Defects — Vacancies and Dislocations

In ionic crystals, where the binding energy is coulombic, an applied electric field tilts the energy level diagram as shown in Fig. 4. Then, the probability of a negative ion acquiring enough thermal energy to surmount the hill on the right is greater than the probability of getting enough energy to surmount the hill on the left. Consequently, if there are vacancies in the negative lattice, negative ions will drift to the right. Similarly, positive ions will drift to the left. These processes constitute electrolysis. The activation energy for electrolysis differs by very little from that for diffusion and in practice is the same.

The evaporation process mentioned above has its counterpart in absorption. This counterpart must exist or all solids would evaporate away in time — some in a very short time. The fact is that solids are in equilibrium with their surroundings. Such an equilibrium is illustrated in Fig. 5. The figure shows schematically a solid having N sites that a particular constituent can occupy, n_S of which are occupied. Also shown is the vapor phase which has N_c sites for the same constituent, of which n_V are occupied. The rate at which particles move from the solid to the vapor phase and from the vapor to the solid phase is discussed in Appendix B; the equilibrium formula is found to be (in the notation of Fig. 5)

$$n_S N_c W_{12} = n_V (N - n_S) W_{21} \tag{7}$$

where W_{12} is the probability that the constituent of concern moves from state 1 to state 2 in unit time (see Appendix B)

$$\frac{W_{21}}{W_{12}} = \exp (\Delta H/kT) \tag{8}$$

and ΔH = the latent heat of sublimation.

It frequently happens that

$$n_S << N \tag{9}$$

then

$$n_V = n_S \frac{N_c}{N} \exp (- \Delta H/kT) \tag{10}$$

Eq. (10) says that the number of particles in the vapor phase is directly proportional to the number in the solid phase and that their ratio is determined very sensitively by the latent heat of sublimation ΔH; conversely, to maintain a certain constituent in a solid phase, a certain vapor pressure of this constituent must be maintained about it. When the latent heat is high, the required pressure may be very small indeed but is still finite.

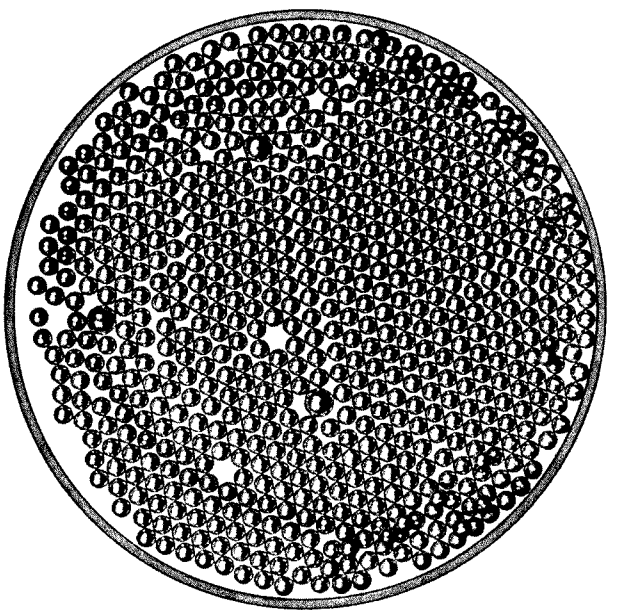


Figure 2. Ball-Bearing Model of a Two-Dimensional Crystal with Impurities (Over-size Balls). (Note that the crystal accommodates the impurities by forming vacancies and dislocations)

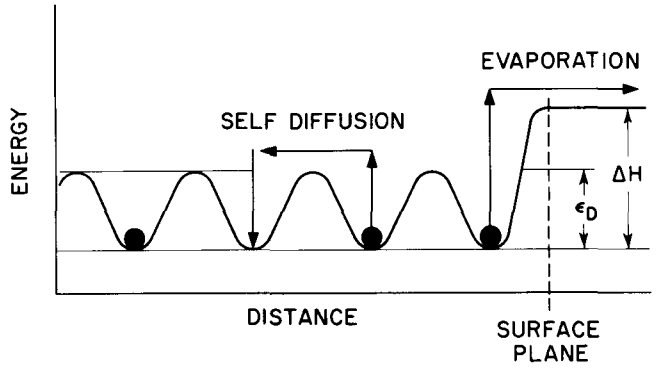


Figure 3. The Energy of an Atom as a Function of Distance Through a Crystal

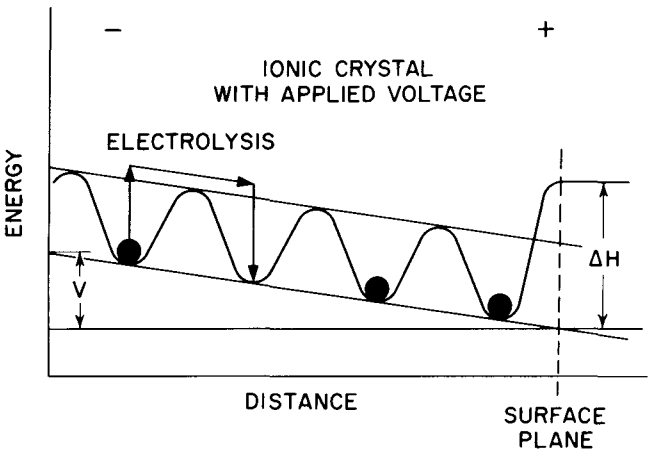


Figure 4. The Periodic Electric Field with a Voltage Applied to an Ionic Crystal

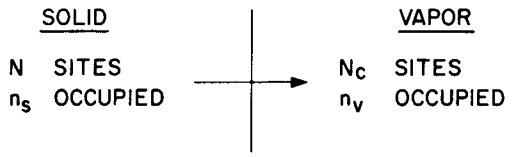


Figure 5. Solid-Vapor Relationships in the Equilibrium Between a Solid and a "Vacuum"

If the ambient vapor is continuously pumped away, the "saturated emission" current is the product of n_v and the mean thermal velocity is normal to the surface. This velocity is given by

$$\bar{v}_n = \frac{1}{4}\bar{v} \qquad (11)$$

where $\bar{v} = \sqrt{\frac{8\,kT}{\pi\,m}}$

m = the mass of the particle

($\bar{v}_n \sim 10^7$ cm sec^{-1} at room temperature)

Thus the particle current density, i.e., the rate of evaporation per square centimeter, is given by

$$j_p = n_v \bar{v}_n \qquad (12)$$

Electronic Properties of Solids

Band Structure. The discussion so far has related to the atomic constituents of solids. It is now pertinent to consider the behavior of electrons within the atoms or ions that constitute the solid.

As has been noted earlier, an atom in free space consists of a compact nucleus carrying an integral number of positive electronic charges surrounded by an equal number of electrons. To free each electron from the nuclear field requires a specific amount of energy. This situation is illustrared in Fig. 6 which shows the energy-level diagram of a valence electron of a simple atom. The energy level of the electron is shown as at 1 in the figure, together with an excitation level for this electron at 1e. To raise the electron into the bottom of the continuum of free energy states, requires the ionization energy ϵ_1. When atoms are compacted into a crystal, the forces between atoms shift the electron energy levels in much the same way that coupling two isochronous tuned circuits changes the resonant frequencies. Just as the resonant frequencies of a pair of coupled circuits must be ascribed to the circuits in combination and not to each circuit individually, the energy levels of the atoms "split" to form a band of permitted levels which must be associated with the crystal as a whole instead of with individual atoms. The "splitting" of the uppermost atomic level and its excitation level may occur in one of two ways depending on the atoms involved. First, the "splitting" may occur in such a way that the resultant bands of levels overlap each other and the free electron continuum above the surface potential barrier or second, the splitting may occur in such a way that the resultant bands of occupied and unoccupied levels do not overlap. Fig. 7 illustrates the ocurrence of such overlapping. In this case, there are about as many free electrons as there are atoms and the material is metallic. Such materials have high conductivity. To be emitted, an electron needs to acquire an energy sufficient to surmount the surface barrier ϕ in Fig. 7, where ϕ is the work function of the metal and is the same as the photoelectric threshold. The case where overlapping does not occur is illustrated in Fig. 8. In this case, there is a region of width ϵ_g which cannot be occupied by electrons; this region is called the "forbidden gap." Materials with a gap between the occupied and free electron bands are semiconductors or insulators. To obtain conductivity, electrons must be thermally excited so that their energy is high enough to enable them to cross the gap into free electron levels.

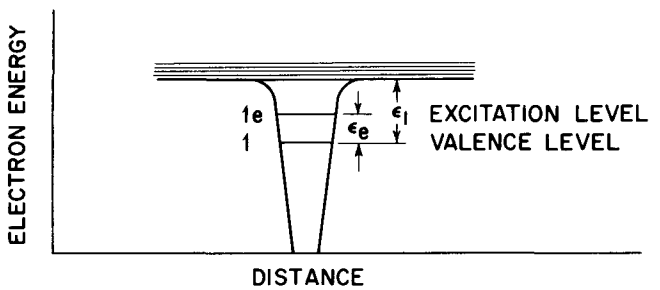


Figure 6. Energy-Level Diagram of a Simple Atom Containing Two Electrons

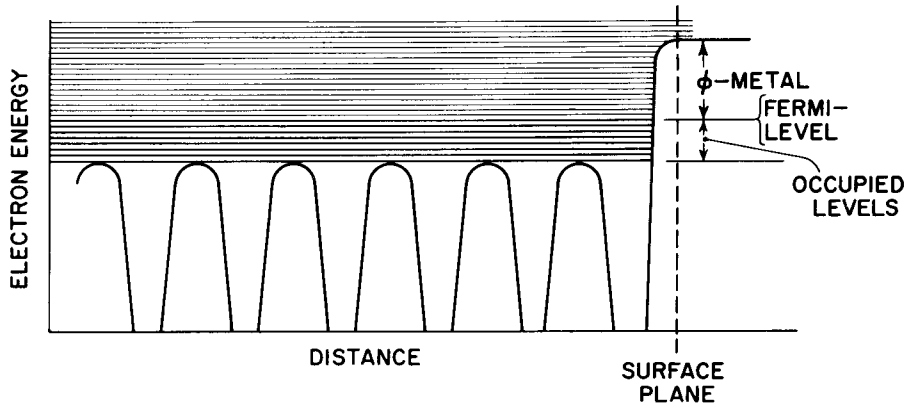


Figure 7. "Splitting" of energy levels of atoms in a crystal to form bands. If the bands overlap, as in this figure, the crystal is metallic.

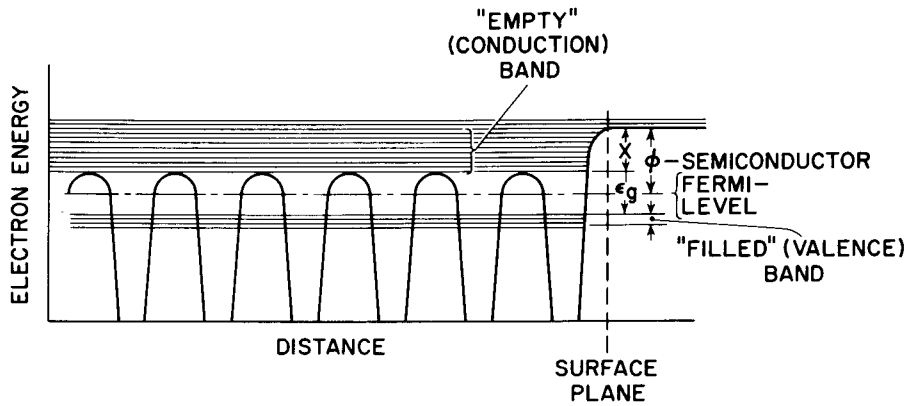


Figure 8. "Splitting" of energy levels of atoms in a crystal to form bands. If the bands do not overlap, a forbidden gap forms and the crystal is, therefore, a semiconductor or insulator.

Except at absolute zero temperature, where the filled band is actually full and the free states are empty, some electrons from the "filled" band are always in the "empty" states. To avoid this contradictory nomenclature, it is advisable to introduce other terms for these bands at this point. Because the filled band is occupied by valence electrons, it is called the valence band. Because electrons in the free states can conduct, this band of states is called the conduction band. It now remains to compute how many electrons from the valence band are to be found in the conduction band at any temperature. The conduction band has N_c states that may be occupied. Because the density of states in the valence band is so high, it may be regarded as also having N_c states which may be occupied. Suppose n electrons from the valence band are in the conduction band. The equilibrium number of such electrons is given by

$$(N_c - n)^2 \; W_{12} = n^2 \; W_{21} \qquad (13)$$

Because $n \ll N_c$ for all reasonable temperatures

$$n = N_c \left(\frac{W_{12}}{W_{21}}\right)^{\frac{1}{2}} = N_c \exp(-\epsilon_g/2kT) \qquad (14)$$

where ϵ_g is the band-gap energy

Therefore, the activation energy for the thermal excitation of electrons into the conduction band is one-half of the bandgap energy. Because there are as many missing electrons ("holes") in the filled band as there are electrons in the conduction band, the Fermi level lies half-way between bands. The semiconductor or insulator is then said to be "intrinsic." To evaporate, an electron in the free energy states (the conduction band) must acquire an additional energy X. This energy is equal to the energy difference between the lowest conduction band levels (the only ones occupied at reasonable temperatures) and the surface barrier. Thus, the thermionic work function is $X + \epsilon_g/2$.

Conductivity. If there are no electrons in the conduction band of a solid, or no holes in the valence band, the solid is non-conducting. The first part of the foregoing statement is obvious, the latter may require explanation. If all of the levels of a band, separated from other bands by substantially more than the thermal energy of the electrons, are occupied, no electron in the band can be accelerated. Acceleration would result in an increase in energy, and there are no empty energy states into which the electron can move. Hence, solids with large band-gaps are insulators. Electrons in the conduction band of a solid have many unoccupied states

into which they can move, so they can be accelerated. Similarly, holes in the valence band open up energy states into which electrons can move, thus permitting acceleration. Whenever an electron moves into an unoccupied state in the valence band, it leaves a "hole" behind just as do atoms in vacancy diffusion. It is convenient to think of the hole rather than the electrons as moving. The hole then acts like a positive charge. In summary, both conduction electrons and valence-band holes can act as charge carriers; the former are negative (n-type), the latter are positive (p-type).

Both electrons and holes move with thermal velocities, velocities of the order of 10^7 cm per sec. In so doing, they collide with lattice atoms or impurities now and then. If the average distance between collisions is L, the average time between collisions is

$$\tau = \frac{L}{\bar{v}} \tag{15}$$

where $\bar{v}$ is the thermal velocity, as previously defined

In the presence of an electric field E, an electron or hole, starting from rest, moves a distance $s = (q/2m)Et$ in a time t. When $t = \tau$ the electron makes a collision and must start from rest again.

Hence

$$L = \frac{1}{2}\frac{q}{m}E\,\tau^2 \tag{16}$$

And the average drift velocity due to the field is

$$v = \frac{L}{\tau} = \frac{1}{2}\frac{q}{m}E\,\tau = \frac{1}{2}\frac{q}{m}\frac{L}{\bar{v}}E \tag{17}$$

This equation is not exact because some electrons have a remanent forward component of velocity after a collision and others moving at angles to the field are scattered forward. A more refined computation taking these effects into account leads to

$$v = \frac{8}{\pi}\frac{q}{m}\frac{L}{\bar{v}}E \tag{18}$$

instead of Eq. (17).

In both computations, it has been assumed that v is small compared to $\bar{v}$ so that the collision time is determined by the thermal velocity. This assumption is valid for all the usual applied voltages. The ratio of drift velocity to field is called the mobility μ, where

$$\mu = \frac{8}{\pi}\frac{q}{m}\frac{L}{\bar{v}} \tag{19}$$

where Eq. (18) has been used instead of the less exact Eq. (17)

Note that in many solids, the apparent mass of electrons and holes is not the free-space mass of electrons. Due to the periodic fields they see in crystals, their masses may be as small as a few per cent of the free-electron mass and theoretically may become negative in certain instances. These apparent or effective masses, as they are called, are usually denoted by m^*.

The conductivity, then, is the product of the number of carriers, the charge they carry, and the mobility, or

$$\sigma = nq\mu \tag{20}$$

Diffusion. The diffusion of electrons or holes — or atoms, ions, or molecules, for that matter -- occurs when their density varies from point to point. This variation in density gives rise to a current in the absence of an electric field and is important in many semiconductor devices, including transistors. To illustrate the formation of such a current, consider a plane normal to a one-dimensional distribution of electrons, as shown in Fig. 9. All of the electrons with velocity directed to the right, in a column of length L and unit-area cross section on the left of the reference plane, will, on the average, cross the reference plane in a time τ without making a collision. Similarly, the electrons with velocity directed to the left in a similar column on the right will move to the left. The average density of electrons on the left is $\left[n - (dn/dx)(L/2)\right]$; the density on the right is $\left[n + (dn/dx)(L/2)\right]$.

Hence, the net rate of flow to the right is

$$j_D = \left(n - \frac{dn}{dx}\frac{L}{2}\right)\frac{L}{\tau} - \left(n + \frac{dn}{dx}\frac{L}{2}\right)\frac{L}{\tau} = -\frac{L^2}{\tau}\frac{dn}{dx} \tag{21}$$

or

$$j_D = -\bar{v}\,L\frac{dn}{dx} \tag{22}$$

Now j_D is the diffusion current density and, by definition, the diffusion constant is

$$D = \bar{v}\,L \tag{23}$$

The relation between the mobility and the diffusion constant is

$$\frac{\mu}{D} = \frac{8}{\pi}\frac{qL}{m\bar{v}}\frac{1}{\bar{v}L} = \frac{q}{kT} \tag{24}$$

This expression is the Einstein relation.

In a region where positive charges, such as holes, are non-uniformly distributed in the x-direction, and a field is applied in the x-direction the net current density is

$$j = ne\,\mu E - De\frac{dn}{dx} \tag{25}$$

If the current is made zero, the equation is easily integrable because $-\int E\,dx = V$, the voltage drop.

The result is

$$n = n_0 \exp(-eV/kT) \tag{26}$$

where n_0 is the density when $V = 0$.

This expression is the well-known Boltzmann equation.

Donors, Acceptors, and Traps. So far, it has been assumed that the crystals have been free from defects and have contained no foreign atoms. If this assumption were true, semiconductors and insulators would have little utility — transistors would be impossible

and the oxide cathode would not exist. Fortunately, it is possible to introduce foreign atoms, ions, and vacancies into solids to make them useful. In a covalent crystal, the substitution of a fraction of atoms having one more electron than the host-crystal atoms provides electron "donors." These atoms produce energy levels that lie in the forbidden gap near the conduction band, and are therefore easily ionized. A conventional energy-level diagram with donor levels is shown in Fig. 10. In this figure, the horizontal lines of Fig. 8, indicating the energy levels that constitute the valence and conduction bands, are omitted and only the uppermost state of the valence band and the lowest state of the conduction band are shown. The potential barrier at the surface is indicated by a step. The energy levels introduced by the donors are shown by the dotted horizontal line, dotted to indicate that the donors are localized and do not give rise to a band that extends throughout the crystal. Similarly, the substitution of atoms having one less valence electron than the host atoms of a covalent crystal produces "acceptor levels" near the valence band, levels which by accepting electrons from the valence band thus produce holes in the valence band. Donors can be produced in ionic crystals by substituting ions of higher valence than the negative constituent ions of the crystal, or of lower valence than the positive constituent ions of the crystal. Conversely, acceptors can be produced by substituting ions of lower valence than the negative constituent ions of the crystal, or of higher valence than the positive constituent ions. Vacancies in the negative lattice of ionic crystals also provide donors; such donors are known as F-centers. Positive ion vacancies yield acceptors and are known as V-centers. To maintain charge neutrality of the crystal, the vacancies must be occupied by as many electrons as constituted the net negative charge of the ion removed.

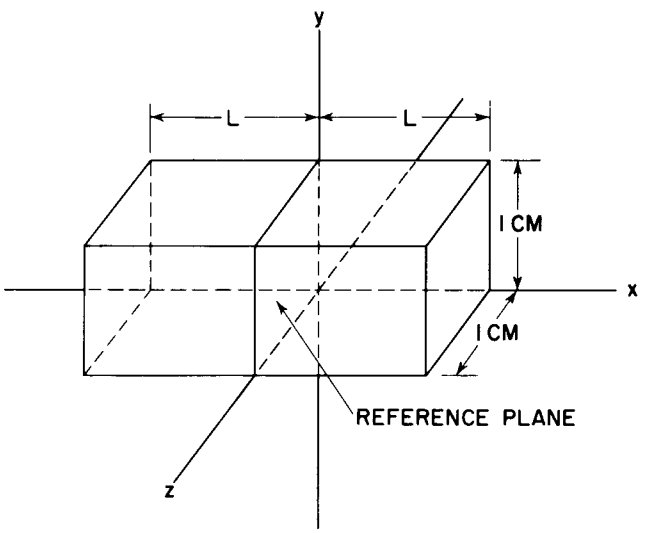


Figure 9. Column for the Computation of a Diffusion Current

Because donor centers and acceptor centers lie in the forbidden gap, they require less thermal energy for ionization than does excitation from the valence band to the conduction band.* The number of electrons excited to the conduction band from donors of density N_D lying at an energy ϵ_D below the conduction band is easily computed, by use of the by-now familiar formalism for computing equilibrium distributions. The resulting equation is

$$(N_D - n)\ N_c\, W_{12} = n^2 W_{21} \tag{27}$$

where n is the density of ionized donors, i.e. the number of electrons in the conduction band.

The solution of this quadratic equation is

$$n = \frac{Nc}{2} \exp(-\epsilon_D/kT) \sqrt{1 + \frac{4N_D}{N_c} \exp(\epsilon_D/kT) - 1} \tag{28}$$

When

$$N_D \ll N_c \exp(-\epsilon_D/kT) \tag{29}$$

All of the donors are ionized, and

$$n = N_D \tag{30}$$

When

$$N_D \gg N_c \exp(-\epsilon_D/kT) \tag{31}$$

Only a fraction of the donors are ionized, and

$$n = \sqrt{N_D N_c} \exp(-\epsilon_D/kT) \tag{32}$$

and the activation energy is one-half the difference in energy between the donors and the edge of the conduction band. Fig. 11 is a plot of Eq. (28) for the particular case $(N_c/2)\exp(-\epsilon_D/kT) = 10^{10}$ and shows the change from full ionization to partial ionization as the density of donors is increased.

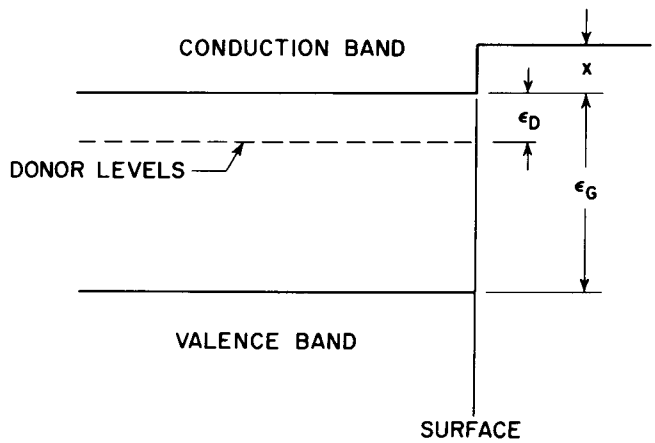


Figure 10. Conventional Energy-Level Diagram for a Solid

*It is assumed that the energy levels of other donors and of the valence band lie far enough below the energy levels of the donors considered so that these "other donors" and the valence band do not contribute significantly to the occupation of the conduction band.

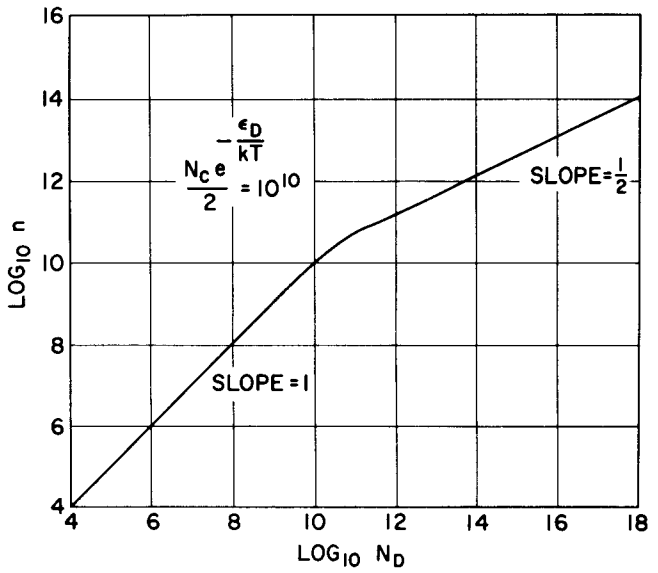


Figure 11. The Density of Electrons n in the Conduction Band of a Solid as a Function of the Donor Density N_D When $N_c e^{\epsilon_D/kT} = 2 \times 10^{10}$

The "doping" atoms normally used in germanium and silicon have such low activation energies that Eq. (29) is satisfied, and all of the donors and acceptors are ionized at room temperature. However, the donors in BaO lie far enough below the conduction band so that Eq. (31) is satisfied, even at the temperatures at which oxide cathodes are operated (1000 to 1100 K). As a result, the oxide cathode requires a high density of donors to provide useful conductivity and emission. These requirements will be discussed in detail later.

Donor and acceptor levels are not the only levels produced by foreign atoms, introduced by intent or accidentally. Unfortunately, foreign atoms may produce levels known as "trapping levels." These are levels in the forbidden gap, unoccupied at zero temperature, which drain electrons from the conduction band, either directly or by emptying donors lying above them. They are particularly pernicious in photoconducting devices. When electrons are excited from the valence band to the conduction band by incident light, many are trapped by these levels. This trapping reduces the photoconductivity and, what is worse, these traps empty very slowly so that electrons dribble back into the conduction band from the traps for a long time after the light has been shut off. This dribbling back results in a very slow decay in the photoconductivity. At present, trapping levels set a definite (and embarrassingly low) limit to the speed of response of photoconductive devices.

Other foreign atoms form recombination centers. These centers trap electrons (or holes). Having done so, they then trap holes (or electrons) which recombine with the charge already trapped, thus destroying electrons and holes in pairs. Such centers may limit the equilibrium densities of electrons and holes obtainable and, therefore, the recombination through centers rather than direct recombination may limit conductivity.

Conclusion

This section has skimmed the aspects of solid-state physics that pertain to electron emission; the viewpoint has been elementary, some would say naive. However, the intent has been to outline the basic phenomena involved in emission problems and to introduce the jargon that will be encountered in the literature on the subject, without at the same time burdening the reader with mathematical rigor.

EMISSION MECHANISMS

This section will be devoted to an outline of the various schemes by which electrons are persuaded to leave their natural habitat in solids so that they may be exploited in an environment free from the lattice and impurity collisions that impede their motions in solids. Only the basic mechanisms will be considered and no attempt will be made to detail the artifices used to exploit the various mechanisms.

Thermionic Emission

Thermionic emission is simply the thermal evaporation of electrons from a solid. Hence, Eq. (12), derived in the preceding section, may be used to write down the evaporation current density

$$j_p = n_v \bar{v}_n \tag{12}$$

This equation gives the particle current. To obtain the electron current, Eq. (12) must be multiplied by the electron charge q, thus

$$j_e = n_v q \bar{v}_n \tag{33}$$

For a semiconductor, the number of electrons N_v in the vapor phase is related to the number of electrons in the conduction band by Eq. (10)

$$n_v = n_s \frac{N_c}{N} \exp(-\Delta H/kT) \tag{10}$$

For electrons, the latent heat ΔH is the product of the electron charge q and the electron affinity X^*; the number of possible sites N for electrons in the conduction band is N_c; and the number of occupied sites n_s is the number of electrons in the conduction band n. Hence, Eq. (10) reduces to

$$n_v = n \exp(-qX/kT) \tag{34}$$

To obtain the emission formula, it remains to express n in terms of donor density in the semiconductor. Eq. (32) gives the required relation for the semiconductors used as thermionic emitters

$$n = \sqrt{N_D N_c} \exp(-\epsilon_D/2kT) \tag{32}$$

Combining Eq. (33), (34), and (32)

$$j_e = e\bar{v}_n \sqrt{N_c N_D} \exp\left[\frac{-q(X + \epsilon_D/2)}{kT}\right]$$

or

*In emission problems, it is customary to express X, and ϵ_D, and ϕ in terms of electric potential, i.e., electron volts; this convention will be used henceforth.

$$j_e = AT^2\sqrt{\frac{N_D}{N_c}}\ \exp\left[\frac{-q\ (X + \epsilon_D/2)}{kT}\right] \qquad (35)$$

where

$$A = \frac{4\pi k^2\ mq}{h^3} \qquad (36)$$

In practical units, the value of A is the familiar

$$A = 120 \text{ amperes cm}^{-2} \text{ deg}^{-2}$$

Note that the work function is the sum of the electron affinity X and one-half the energy gap between the donors and the edge of the conduction band.

For metals where the valence band overlaps the conduction band (see Fig. 7), the solid may be regarded as an intrinsic semiconductor (one having no donors, acceptors or traps) with zero bandgap. Then, Eq. (14) with ϵ_g set equal to zero is applicable, i.e., $n = N_c$. Using this value for the electron density in the conduction band is tantamount to setting $N_D = N_c$, and $\epsilon_D = 0$ in Eq. (35).

Thus for a metal

$$j_e = AT^2\ \exp\ (-q\phi/kT) \qquad (37)$$

where X of Eq. (35) has been replaced by ϕ, the usual symbol for the work function of a metal.

The work functions of metals are usually high (of the order of 4.5 electron volts) so they are ordinarily used only in high-power transmitting tubes where their thermal stability and insensitiveness to ambient gases make them desirable.

The work function of a metal can be reduced by building a dipole layer on the cathode surface. Fig. 12 is an energy-level diagram showing a dipole layer consisting of a monolayer of oxygen, on the metal surface, covered by a monolayer of some low-work-function metal, such as cesium or barium. The oxygen underlayer binds the low-work-function metal to the base metal so that the evaporation rate of the low-work-function metal is greatly reduced. Unfortunately, the evaporation rate is still appreciable. Furthermore, the dipole layer is easily removed by ion bombardment, so that any gas in the tube has a serious effect on such an emitter. To mitigate these difficulties, A. W. Hull invented the "dispenser cathode" in which a reservoir of active material replenishes the dipole layer as it is lost by evaporation or bombardment. More recently the dispenser cathode has received much attention. The Lemmens ("L") and Katz ("M-K") cathodes are examples of the dispenser cathode. In some of these cathodes, alkaline earth oxides and a reducing agent are stored in a reservoir. The reducing agent reacts with the oxides to produce free metals which proceed to the emitting surface through a porous plug by Knudsen flow through the pores or by surface diffusion. If the reducing agent reacts rapidly, the porosity of the plug controls the rate of replenishment; if the reducing agent reacts slowly, the rate of reaction controls the rate of replenishment. In other versions, alkaline earth compounds (oxides, silicates, and aluminates are used) are compacted with finely divided metal powder to form the emitter. These cathodes are mechanically stable and can be made with precision. Because any loss of dipole layer is replenished, they can be operated at temperatures at which they yield emissions of the order of amperes per cm^2.

The measurement of the work function of an emitter is by no means simple. Part of the difficulty arises from the fact that the work function is lowered by an electron-accelerating field at the emitter surface. This effect is known as the Schottky effect.[1] It arises because an electron leaving the surface sees its image in the surface. Thus, in the case of a metal, the force on an electron is

$$F = \frac{q^2}{(2x)^2} - qE \qquad (38)$$

where x is the electron's distance from the surface, and E is the applied field. At some distance l from the surface, the force becomes zero and the electron can be regarded as free. This distance is obtained by setting F = 0. Thus,

$$l = \frac{1}{2}\sqrt{\frac{q}{E}} \qquad (39)$$

The work required to remove an electron from the surface is

$$W = \int_0^l \left[\frac{q^2}{(2x)^2} - qE\right] dx$$

Examination of this equation shows that something is wrong because the integral is infinite. The difficulty is that the image force cannot vary as x^{-2} down to atomic distances; somewhere near the surface, the image force must change in a manner which keeps the force finite. Schottky ingeniously bypassed the difficulty by dividing up the integral as follows

$$W = \int_0^\infty q f\ (x)\ dx - \int_l^\infty \frac{q^2}{(2x)^2}\ dx - \int_0^l qE\ dx$$

where f (x) represents the image force near the surface and behaves as $q/(2x)^2$ remote from the surface. Integration then gives

$$W = \int_0^\infty qf\ (x)\, dx - \frac{q^2}{4l} - qEl \qquad (40)$$

Eliminating l by use of Eq. (39) gives

$$W = \int_0^\infty qf\,(x)\, dx = q\sqrt{qE}$$

In the absence of field, W must be just $q\phi$, where ϕ is the work function. Therefore,

$$W = q\left[\phi - \sqrt{qE}\right] \qquad (41)$$

In practical units, this relation is

$$W = q\left[\phi - 3.72 \times 10^{-4}\sqrt{E}\right]$$

where E is in volts cm^{-1}

The result shows that the work function is reduced by an electric field and that the equation for the emission of a metal in the presence of a field should be

$$j_e = AT^2 \exp\left[-\frac{q}{kT}\left(\phi - \sqrt{qE}\right)\right] \quad (42)$$

This equation may be written

$$j_e = j_o \exp\left[\frac{q}{kT}\sqrt{qE}\right] \quad (43)$$

where j_o is known as the field-free emission

A plot of the logarithm of j_e versus the square root of the field should yield a straight line with a zero-field intercept of $\ln j_o$. To determine the work function of a metal, it is the logarithm of $j_o T^{-2}$ that should be plotted against T^{-1} to make a so-called Richardson plot. Then, the slope of the line will be ϕ and the intercept will be A.

For semiconductors, the simple image-force representation is not adequate because the density of electrons is usually so low that the field of the external electron extends a considerable distance into the solid — a Debye length. [2] To complicate matters further, the dielectric constant of semiconductors usually differs substantially from unity. The treatment of a problem of this complexity is beyond the scope of a simple exposition such as this. It will only be noted that Schottky plots and Richardson plots made from the zero-current intercepts obtained on semi-conducting cathodes must be treated with great reserve.

When data for Schottky plots are taken with great precision, it is found that the resulting curve shows periodic deviations from the straight line computed above. [3, 4] These deviations are due to the wavelike nature of the electrons. As the field at the surface increases, the potential barrier keeping electrons in the solid becomes narrower and it acts like a "quarter-wave plate" in optics so that, depending on their velocity (wave length), some electrons are more easily transmitted and others are more completely reflected. The net effect of summing over-all velocities is the periodic variation in emission. The effect is very small, but in the hands of Coomes and his co-workers it has provided a powerful tool for studying the nature of surfaces.

Because of the great difficulty in determining true work functions, W. S. Nottingham has suggested that the Richardson equation

$$j_e = AT^2 \exp(-q\phi/kT) \quad (37)$$

be written for all thermionic emitters and that A and ϕ be starred (A^*, ϕ^*) to indicate that they are not true values of the thermionic constants. He proposed the name "work factor" for ϕ^*. This proposal has the merit of providing a basis for the comparison of cathodes without raising the many questions involved in interpreting the constants in fundamental terms.

Field Emission [5, 6]

Field emission results from the wavelike nature of electrons. It is the periodic deviation from the Schottky line effect carried to the extreme where most of emitted electrons "tunnel" through, rather than moving over the potential barrier. Because the computation of the field-emission formula is complex and laborious, the derivation will not be reproduced here. A computation that illuminates the physical mechanism involved and displays the general functional form of the emission equation will suffice.

Fig. 13 is an electron-energy diagram of a metal with a strong field at the surface. As before, ϕ is the work function and η is the width of the occupied levels. The potential outside of the surface is

$$V = \phi - Ex \quad (44)$$

where E is the electric field

The triangular area labeled A is the potential barrier through which the field-emission electrons must penetrate. Also shown is the curve of the distribution of electrons in the various energy states. Notice that there are practically no electrons where the barrier is thin; therefore, most of the emitted electrons must penetrate the barrier where its thickness is

$$S = \frac{\phi}{E} \quad (45)$$

The equation of an electron wave (as in Appendix A) is

$$\psi = \exp\left(\pm i\,\frac{2\pi x}{\lambda}\right) \quad (46)$$

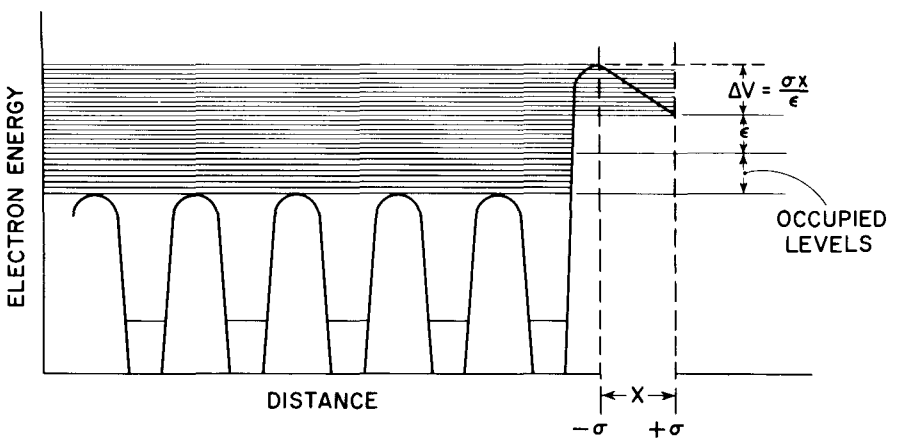


Figure 12. Energy-Level Diagram Showing Reduction of the Work Function of a Metal by Building a Dipole Layer on the Cathode Surface

where

$$\frac{2\pi}{\lambda} = \frac{p}{\hbar} \qquad \text{(See App. A, Eq. 4)}$$

and the momentum p is $\sqrt{2m(\epsilon - U)}$

The total energy ϵ of the electrons at top of the occupied levels (Fermi Sea) is zero.

The average potential energy through the potential barrier is

$$U = \frac{p\phi}{2} \tag{47}$$

Therefore, the momentum is

$$p = \sqrt{-mq\phi} = i\sqrt{mq\phi} \tag{48}$$

And the propagation constant is

$$\frac{2\pi}{\lambda} = i\sqrt{\frac{mq\phi}{\hbar}} \tag{49}$$

Hence, the wave function is

$$\psi = \exp\left(\mp\frac{\sqrt{mq\phi}}{\hbar}x\right) \tag{50}$$

The electron density is a maximum on the left and drops off to the right, so it is clear that the negative sign is pertinent to this case. Thus

$$\psi = \exp\left(-\frac{\sqrt{mq\phi}}{\hbar}x\right) \tag{51}$$

Hence, the density of "free" electrons at x = s is

$$\psi\psi^* = \exp\left(-2\frac{\sqrt{mq\phi}}{\hbar}\frac{\phi}{E}\right) \equiv \exp\left(-C\frac{\phi^{3/2}}{E}\right) \tag{52}$$

This calculation has been very rough; it yields the form of the exponential factor in the emission equation and that is about all. The complete emission formula in practical units is

$$J = 6.2\times10^{6}\left(\frac{\eta}{\phi}\right)^{1/2}\frac{E^2}{\eta+\phi}\exp\left(-6.84\times10^{7}\frac{\phi^{3/2}}{E}\right) \tag{53}$$

For the refractory metals, ϕ is about 4.5 electron volts and η is about 10 volts. With ϕ = 4.5 electron volts, the exponential is approximately

$$e^{-\left(\frac{6.5\times10^{8}}{E}\right)} \sim 10^{-\left(\frac{2.8\times10^{8}}{E}\right)}$$

Hence, a field of about 1.3×10^7 volts per centimeter is required to obtain an emission current of an ampere per square centimeter. Because the emission varies so rapidly with field, a very slight increase in field will run the current into thousands of amperes per square centimeter. However, such field strengths can be obtained only with very sharp points and the heating of these points at the high current densities obtained gives rise to surface diffusion of the atoms of the metal. The surface diffusion tends to lower the surface energy of the solid and rounds off the point so that the field decreases, and with it, the emission. Hence, only the most refractory metals have found use as field emitters and these, chiefly in field-emission microscopes. An attempt has been made to use germanium as a field-emitter in a surface study; however, the point deformed badly and, on the whole, little was learned.

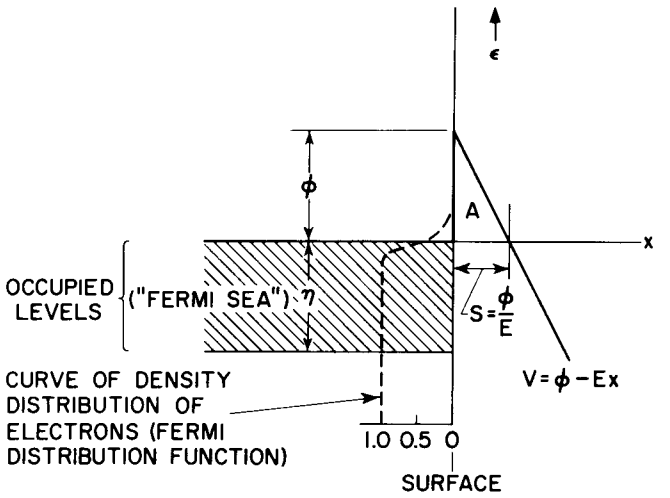


Figure 13. Energy-Level Diagram of a Metal with an Intense Electric Field at the Surface so that Field Emission Occurs

Photoemission[7, 8]

Photoemission is the ejection of electrons from a solid by photons. The first question to be answered is, what electrons? A consideration of the results of a head-on collision between a photon and a free electron will elucidate the problem. The energy transferred from a particle of mass m_1 and energy ϵ_1 to a second particle of mass m_2, at rest initially (keeping in mind conservation of energy and momentum) is

$$\epsilon_2 = \epsilon_1\frac{4m_1m_2}{(m_1+m_2)^2} \tag{54}$$

Suppose the first particle is a photon and the second is an electron. The phenomenon involved is Compton scattering, but to avoid the complications of a relativistic computation a heuristic approach which leads to the correct answer will be used.

As noted earlier, the momentum of a photon is

$$p = \frac{h\nu}{c} \tag{3}$$

If the energy $h\nu$ is written as

$$h\nu = \frac{p^2}{2m^*} \tag{55}$$

where p is the momentum and m^* is the apparent mass

The apparent mass is

$$m^* = \frac{h\nu}{2C^2} \tag{56}$$

This mass is very small; if the energy is one electron volt, the mass is $m^* \sim 10^{-33}$ gram.
Because this mass is small compared to the mass of

an electron, Eq. (54) may be written

$$\epsilon_2 = \epsilon_1 4 \frac{m_1}{m_2} = \epsilon_1 4 \frac{m^*}{m_2} \qquad (57)$$

or

$$\epsilon_2 = \epsilon_1 2 \frac{h\nu}{m_2 c^2} \qquad (58)$$

This is the maximum energy transfer in a Compton encounter. If $h\nu$ is expressed in electron volts, Eq. (58) becomes

$$\epsilon_2 = 10^{-5} \; h\nu\epsilon_1 \qquad (59)$$

Hence, a photon in the visible region of the spectrum, say a 2-volt photon, can transfer only a 2×10^{-5} fraction of its energy to an electron.* This energy is far from enough to overcome the work function of any material. It now seems clear that a direct encounter between a photon and an electron cannot cause photoemission and that a "three-body" collision is necessary. i.e., the photoelectron must be at least partially "bound" so that neighboring electrons, atoms, or the lattice as a whole can take part, and the stringent momentum conservation requirement encountered in the free-electron collision can thus be relaxed.

In metals, photoelectrons come from the top of the Fermi sea. Thus, the photoelectric threshold occurs at

$$h\nu = q\phi \qquad (60)$$

where ϕ is the work function

When $h\nu$ exceeds $q\phi$, the ejected electrons have a maximum kinetic energy of

$$1/2\, mv^2 = h\nu - q\phi \qquad (61)$$

This is the Einstein equation which, when verified experimentally, clinched the particle concept of light. Because the emitted electrons do not all come from the same depth in the Fermi sea, the curve of yield vs. energy $h\nu$ has a tail that makes the exact determination of ϕ from photoelectric measurements difficult. This difficulty was obviated by R. W. Fowler who computed the "straggling" effect and arrived at a universal curve that makes possible the determination of ϕ

Fowler's formula is given by

$$\ln \frac{I}{T^2} = B + \ln\left[\frac{h(\nu - \nu_0)}{kT}\right] \qquad (62)$$

where

I = the photoelectric current
B = a constant
T = the emitter temperature
h = Planck's constant
k = Boltzmann's constant
ν = the frequency of the incident light
ν_0 = the frequency of the light when $h\nu = q\phi$

and

$$f(x) = \begin{cases} e^x - \frac{e^{2x}}{2^2} + \frac{e^{3x}}{3^2} + \cdots, & x \leq 0 \\ \frac{\pi^2}{6} + \frac{x^2}{2} - \left(e^{-x} - \frac{e^{-2x}}{2^2} + \frac{e^{-3x}}{3^2} + \cdots\right), & x \geq 0 \end{cases}$$

The use of the so-called Fowler plot is shown in Fig. 14. The curves labeled T_1 and T_2 are "experimental" curves taken at temperatures T_1 and T_2, respectively. These are matched to the universal curve by translating them horizontally and vertically. The required horizontal translation gives $q\phi/kT$ from which ϕ may be computed; the required vertical translation gives the constant B.

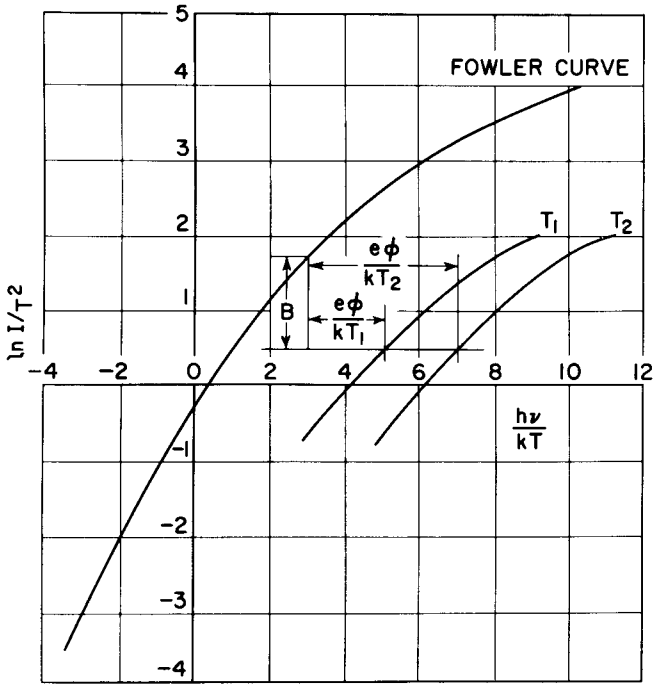


Figure 14. The "Fowler Curve" for Photoelectric Emission and an Example of How the Work Function of a Metal is Determined by Matching Experimental Curves to the Fowler Curve

In semiconductors and insulators, photoelectrons may be obtained from the donor levels and from the valence band. In the former case, a peak, centered at $\epsilon_D + X$, is found when the yield is plotted against the energy of the incident photon.* In the latter case, the yield increases for photon energies exceeding the threshold value of $\epsilon_g + X$. Again the yield curve has a "tail" near the threshold. Attempts to determine the true threshold by matching the Fowler curve have met with varying degrees of success; it is not yet clear that a curve based on a metallic model will fit a semiconductor.

*Similar considerations apply to "electron sputtering." The ratio of electron-mass to atom-mass is of about the same order of magnitude as that above. It seems likely that in "electron sputtering" the incident electron interacts with another electron in an atom or ion, then excites it to a higher electronic state for which the binding energy of the atom or ion to the host crystal is small with the result that the atom or ion departs via dissociation.

*As before, ϵ_D is the depth of the donors below the edge of the conduction band, X is the electron affinity, and ϵ_g is the band-gap energy.

Fig. 15 shows curves of the photoemission of BaO in the inactive state and in a thermionically active state.[9] The effect of the donors produced by activation is apparent in the substantial "tail" on the "active" curve that extends to much lower photon energies than in the inactive case. This tail may be resolved into peaks indicating a set of donor levels at about 2.3 ev below the conduction band and a second set of donor levels at about 1.4 ev below the conduction band. For comparison, Fig. 16 shows the corresponding curves of photoconduction. These curves are displaced from the photoemission curves by $h\nu \sim 0.5$ ev. In photoconduction electrons need only be excited to the conduction band whereas in photoemission they must be excited over the surface barrier. Hence, the displacement between the photoemission and photoconductivity curves must be equal to the electron affinity, i. e., $X \sim 0.5$ ev in the case of BaO.

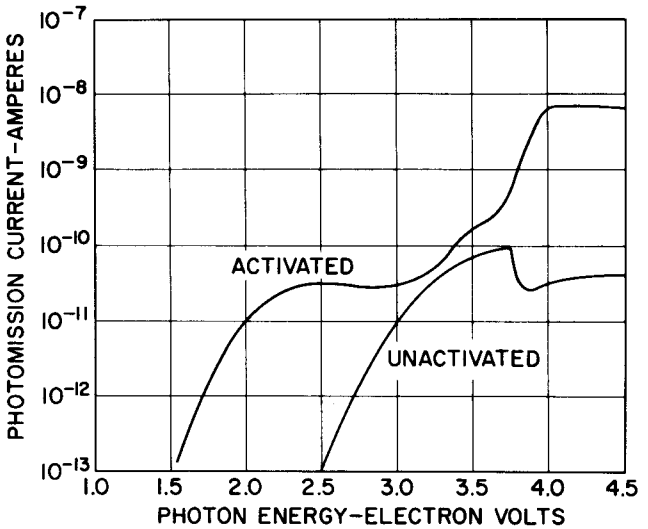


Figure 15. The Photoemission Current vs. Photon Energy for Activated and Unactivated BaO

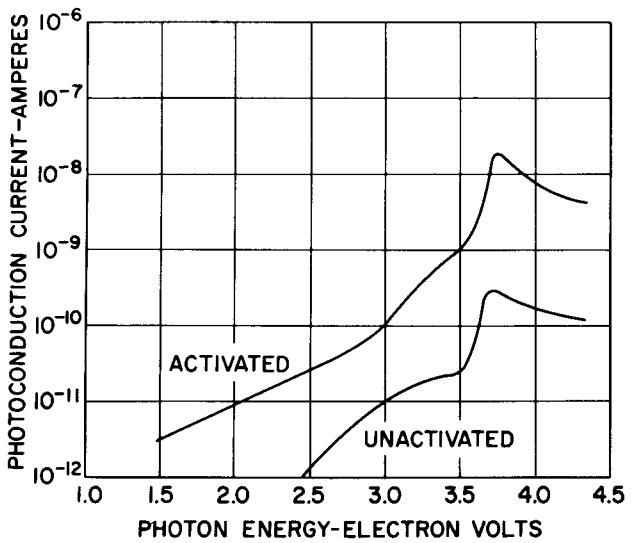


Figure 16. The Photoconduction Current vs. Photon Energy for Activated and Unactivated BaO

The photoelectric yield of metals may be quite high. Hughes and DuBridge cite values of 1 electron for every 14 photons for potassium, and 1 electron for every 4 photons for gold. The usual yield in television pick-up tubes is about 1 electron for every 100 photons.

Secondary Emission[10, 11]

Secondary emission is the ejection of electrons from a solid by electrons impinging on the solid. The theory of secondary emission is in an unhappy state – no theory fits the experimental results in detail. To illustrate some of the physical notions involved, and to point up some of the features common to all present theories, the theory of Bruining will be discussed.

The first problem is that of computing the energy loss of the impinging electron as it moves into the solid. At the energies required for secondary emission, the electron wave length is so small that the chance of a head-on collision with any other electron in the solid is trivial. Hence the encounter of the oncoming electron with any other electron is fleeting. The momentum transferred to the secondary electron may be written

$$\Delta p = \bar{F} t \tag{63}$$

where $\bar{F}$ is the average force felt during the encounter time t, in other words $\bar{F}t$ is the impulse.

Now t depends on the range l of the force $\bar{F}$ and the velocity of the primary electron

$$v = \sqrt{\frac{2qV}{m}} \tag{64}$$

where V is the energy in electron volts.

Therefore

$$\Delta p = \frac{\bar{F} l}{\sqrt{\frac{2qV}{m}}} \tag{65}$$

The energy imparted to the secondary electron is

$$\Delta \epsilon = \frac{(\Delta p)^2}{2m} = \frac{(\bar{F} l)^2}{2m\frac{(2qV)}{m}} \tag{66}$$

This is also the loss of energy of the primary electron. If there are n encounters per centimeter of path, the energy of the primary electron decreases as

$$\frac{\partial \epsilon}{\partial x} = - n \frac{(\bar{F} l)^2}{2m\left(\frac{2qV}{m}\right)} \tag{67}$$

If the energy ϵ is expressed in terms of V

$$\frac{\partial V}{\partial x} = - \frac{n (\bar{F} l)^2}{2mq\left(\frac{2qV}{m}\right)} \equiv - \frac{a}{2V} \tag{68}$$

Integrating yields

$$V^2 - V_o^2 = - ax \tag{69}$$

or

$$V = \sqrt{V_o^2 \quad -ax} \tag{70}$$

where V_o is the energy of the electron at the entrance surface of the solid

This equation is the Whiddington Law for the loss of energy of an electron moving into a solid.

Terrill has found that for many metals

$$\frac{a}{\gamma} \sim 0.4 \times 10^{12} \text{ volts}^2 \text{ cm}^{-1} \tag{71}$$

where γ is the density of the metal relative to the density of water.

Bruining then writes the contribution to the secondary-emission current di_s of an element dx at a distance x below the surface as

$$di_s = \beta i_r e^{-\alpha x} \frac{d(qV)}{dx} dx \tag{72}$$

where

β = a constant

i_r = the primary constant

The "Attenuation factor" for secondary electrons is $e^{-\alpha x}$ This attenuation factor accounts for the fact that some of the secondary electrons moving toward the surface are scattered and do not reach the surface. *

The energy given to secondaries in the interval dx is $\frac{\partial (qV)}{\partial x} dx$. The total secondary current is then

$$i_s = \beta i_r \int_o^{x = V_o^2/a} e^{-\alpha x} \frac{\partial (qV)}{\partial x} dx \tag{73}$$

This equation may be simplified by the substitution

$$y = \sqrt{\frac{\alpha}{a}} V \tag{74}$$

It then becomes

$$i_s = \beta i_r \, q\sqrt{\frac{a}{\alpha}} \, e^{-r^2} \int_o^r e^{y^2} dy \tag{75}$$

where

$$r = \sqrt{\frac{\alpha}{a}} V_o \tag{76}$$

This equation has a maximum for

$$V_o = 0.92\sqrt{\frac{a}{\alpha}} \tag{77}$$

However, like all other theoretical expressions for secondary emission, when the theoretical maximum is fitted to the maximum of an experimental curve, the rest of the curve fails to fit. Therefore, nothing is to be gained by further discussion of the theory.

Experimentally, the energy distribution of secondary electrons is as shown in Fig. 17. Most of the secondary electrons have an energy of less than 10 electron volts, as shown by the peak on the left. The peak on the right, centered on the primary energy, results from reflected electrons.

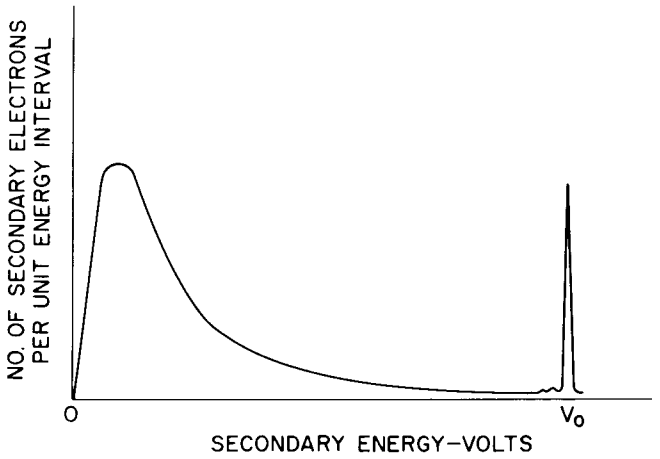


Figure 17. Typical Energy Distribution Curve of Secondary Electrons

The secondary emission ratio δ, the number of secondary electrons divided by the number of primary electrons, of a number of metals is shown approximately in Fig. 18. In every case, the maximum yield occurs at about 600 electron volts.

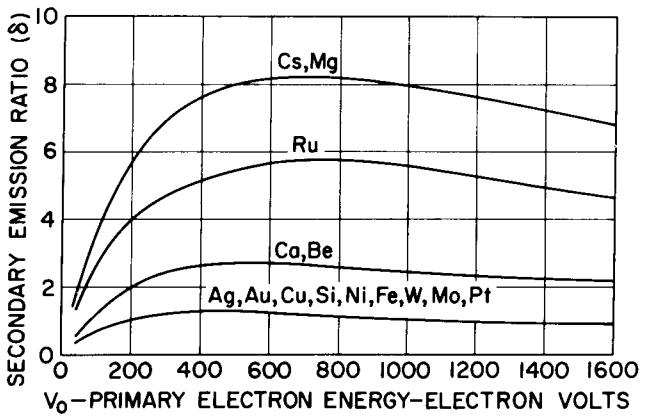


Figure 18. Secondary-Emission Yield vs. Primary-Electron Energy (The curves are approximate and are intended to show only the general behavior and magnitude of the yield of the metals listed)

Compound materials have been found that have yields higher than those of the pure metals. For example, cesium-on-oxygen-on-silver gives a yield of $\delta = 8$ to 11, and magnesium-on-oxygen-on-silver gives a yield of $\delta = 8$. The latter emitter has prompted a study at the University of Minnesota of the secondary-emitting properties of single crystals of magnesium oxide. Recently, Lye at Minnesota has measured the secondary yield from the surface of a single crystal of MgO prepared by cleaving the crystal in a vacuum of 10^{-8} mm of Hg. He obtained an initial yield of $\delta = 24$; the yield dropped with time and electron bombardment.

Some insulating secondary emitters, notably MgO films, display the Malter effect, the emission of elec-

*This factor is used in all secondary-emission theories.

trons for an extended period after the primary beam has been shut off. This phenomenon is believed to result from the charging up of a thin insulating surface layer. Such a charging up produces a field high enough to produce field emission. To illustrate this, suppose that a layer one atom thick were charged to a potential of 1 volt. Then, the field would be approximately 10^8 volts per centimeter. This value is high enough to give a very high initial field-emission current.

It may have been noted that there are some "wrinkles" in the curve immediately to the left of the V_0 peak in Fig. 17. In examining these wrinkles critically and in further experiments on the penetration by electrons of thin metallic films, it has been found that there are discrete energy losses by the primary electrons. Ruthemann and Lang have studied this behavior in detail.[12, 13] Pines and Bohm have advanced the theory that these discrete energy losses result from the collective behavior of the electrons in the metal.[14, 15] The mechanism they envisage is plasma oscillation, a behavior found by Langmuir and Tonks in gas discharges.[1] The computation of the characteristic absorption energy proceeds as follows: Suppose the density of electrons in the Fermi sea is n. Now consider a slab of thickness dx, as in Fig. 19. Suppose the electrons immediately to the left of the slab are displaced to right by an amount ξ and those at the right displaced by an amount $\xi + \frac{d\xi}{dx}$ dx. Then, the net charge moved into the slab is

$$\rho\, dx = qn\,\xi - qn\left(\xi + \frac{d\xi}{dx}\right) = -qn\left(\frac{d\xi}{dx}\right)dx \qquad (78)$$

where ρ is the charge density. Hence, $\rho = -qn\,\partial\xi/\partial x$.

By the Gauss theorem, the divergence of the field E is $\partial E/\partial x = \rho/\epsilon = -(qn/\epsilon)(\partial\xi/\partial x)$, where ϵ here denotes the dielectric constant. Integration yields $E = (-qn/\epsilon)\,\xi$. The integration constant has been set to zero, because the field vanishes when there is no displacement.

The equation of motion of an electron then is

$$\frac{d^2\xi}{dt^2} = \frac{q}{m}E = -\frac{q^2 n}{m}\xi \qquad (79)$$

This equation is that of a simple harmonic motion with an angular frequency

$$\omega_p = \sqrt{\frac{nq^2}{m\epsilon}} \qquad (80)$$

This frequency is the so-called plasma frequency. According to Pines and Bohm, electrons should lose energy in discrete amounts, which Pines calls plasmons, and of magnitude

$$\Delta E = \hbar\,\omega_p \qquad (81)$$

For the case of aluminum, the discrete losses are about 15 electron volts. To see how the theory hangs together, it is interesting to compute the electron density corresponding to this energy loss of 15 electron volts.

Setting

$$E = \hbar\,\omega = qV \qquad (82)$$

with V = 15 qv gives

$$\omega_p \sim 2 \times 10^{16} \qquad (83)$$

From Eq. (80)

$$n = \frac{\omega_p^2\, m\epsilon}{q^2} \sim 10^{23} \qquad (84)$$

This value is of the right order of magnitude for the density of electrons in the conduction band of a metal. This result does not mean that the theory is correct; the fact is that the theory is the subject of much controversy. Many feel that collisions of the electrons with lattice atoms will provide such strong damping that free oscillation cannot take place. Sternglass, who has done much work on the contribution of inner-shell electrons to secondary emission, has suggested that the discrete losses are in fact due to the excitation of electrons from an inner shell and are indicative of the discrete level from which they came.[16] The resolution of this problem is not yet at hand.

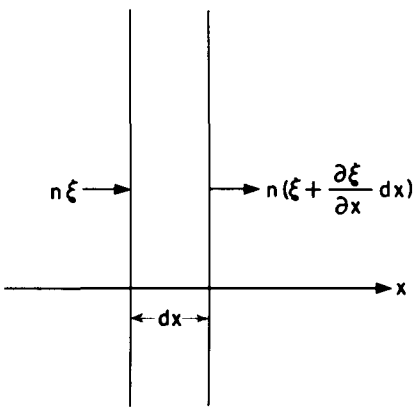


Figure 19. Charge Displacement in an Electron Plasma

THE PHYSICS OF THE OXIDE CATHODE*

Emission-Current Density

The ability of the oxide cathode to emit at current levels of 50-100 amperes cm^{-2} for microsecond periods seemed remarkable at one time. In the light of modern solid-state theory, it does not seem as remarkable. In the previous chapter, the emission equation for a semiconducting cathode was found to be

$$j_e = AT^2\sqrt{\frac{N_D}{N_c}}\,\exp\left[-\frac{q}{kT}\left(X + \frac{\epsilon_D}{2}\right)\right] \qquad (35)$$

*Recent research on oxide cathodes is summarized in a chapter entitled "Electron and Ion Motion" by the present writer, in Halbleiter, III, edited by W. Schottky, Vieweg and Sohn, Braunschweig. This paper contains an extensive bibliography of oxide cathode work, past and present.

At 1000 K, N_c is about 10^{19} for electrons. It should be possible to produce donors to about 1 part in 10^5, i.e., $N_D = 10^{18}$. If these donors lie 1.4 electron volts below the conduction band, as DeVore found in BaO, then, with X = 0.5 electron volts, the current density at 1000K should be

$$j_e \sim 40 \text{ amperes cm}^{-2} \qquad (85)$$

This current density was obtained by Nergaard and Matheson in pulsed diodes.[17] This current density, however, cannot be maintained indefinitely. Ten years ago, the usual explanation of this phenomenon was that the cathode was deactivated by "poisoning." The writer found it difficult to accept this explanation; he had activated many oxide cathodes by drawing current for extended periods, not by shutting it off, yet under pulse conditions the cathode was supposed to be able to reactivate spontaneously in a few milliseconds in the absence of current.

In 1931, Becker and Sears had shown that oxide cathodes could be activated, as least temporarily, by passing current through them in the back direction.[18] They attributed this phenomenon to an electrolysis of what would now be called donors. (They called it Ba.) In 1945, Sproull published the results of his work on the pulse decay of the oxide cathode, and put forward the hypothesis that pulse decay was due to an electrolytic depletion of a dipole layer on the surface of the cathode.[19] These models, involving electrolysis, appealed to the writer and were developed and extended in a series of studies by the writer and his co-workers (R. M. Matheson, R. H. Plumlee, and H. B. DeVore) to give a coherent model of oxide-cathode behavior.

Model of the Oxide Cathode

The model conceives of the oxide cathode as consisting of a semiconducting host material (BaO or a combination of alkaline-earth oxides) containing mobile donors produced by any of a variety of chemical reduction processes. The mobility of the donors permits them to electrolyze and diffuse at the normal operating temperature of the cathode. Not only are the donors mobile, but other constituents of the cathode, such as CO_2, O_2, H_2, H_2O, or their dissociation products, are mobile as well. This may seem to lead to a very complex model of the cathode; however, the model is no more complex than the cathode itself and it is the complexity of the cathode that makes it the ubiquitous emitter that it is.

To give a coherent account of the implications of the mobile-donor model,[20] its principal consequences will be outlined first; then these consequences will be illustrated by experimental data. In brief these consequences are:

1. When no current is flowing, the donors are uniformly distributed throughout the cathode; when current is drawn, the donors electrolyze towards the base metal. This flow sets up a donor density gradient so that back-diffusion of donors occurs. When the current is constant, electrolysis is everywhere balanced by back-diffusion. The donors do not plate out at the base metal, because of their low solubility in the metal.

2. The resistance of the cathode is a function of the average current. The average current determines the distribution of donors which determine the resistivity. Because donors are depleted from the region of the emitting surface, the voltage drop near the surface is high; because donors accumulate towards the base metal, the voltage drop near the base metal is low. The net effect is that the resistance of the coating rises with the current through it.

3. At modest current levels, a diode may show marked departures from the Child-Langmuir law without being emission limited. This departure is due to the cathode-resistance voltage drop, which reduces the voltage across the vacuum diode. The increase of cathode resistance with life may, in fact, determine the useful life of a tube.

4. At very high current levels, donor depletion reduces the electron density at the emitting surface to the point that emission limitation sets in. It then becomes apparent that the work function depends on the average current.

5. If a pulse is applied to a diode, the initial current will be high but, as the donors redistribute and the resistance of the cathode rises, the current will decay correspondingly. The initial current in a second pulse will depend on the extent of the electrolysis during the first pulse (i.e., on the current density and pulse length) and on the time between pulses during which the donors can diffuse towards the zero-current uniform distribution.

6. If donors are produced at the cathode surface during current flow, the donors flow through the surface-depletion layer into the bulk of the cathode and, in so doing, reduce the resistance of the depletion layer and greatly enhance the electron current. Subsequent to donor production, the conductance of the cathode is increased in proportion to the number of donors added.

7. The electrolysis of oxygen ions in the cathodes increases the oxygen-ion density at the surface and thus enhances the evaporation of oxygen. This electrolysis is a reduction process in which the external power supply that produces the current acts as a reducing agent. It is the mechanism that permits the activation of cathodes on passive base metals, and the mechanism by which a cathode maintains its activity for thousands of hours in the presence of residual oxygen in tubes.

8. Other constituents of the cathode also electrolyze. Among those studies are CO_2, H_2, and H_2O. Particularly interesting is the case of H_2O; it is found that H_2O is absorbed by the cathode when oxygen is eliminated. This effect suggests that the donor in the cathode is actually $Ba^{++}(OHe)^{--}$ and that it is the spare electron (e) that is easily ionized to provide conductivity. Then, donor electrolysis occurs by proton (H^+) transfer, and the rapid decay of current

under pulse conditions is easily accounted for.

These items do not exhaust the numerous aspects of the oxide cathode that might be discussed in terms of the present model. However, they illustrate the features pertinent to the activation and life of cathodes and to the limitations of a cathode in practical use.

Cathode Conductivity

In considering cathode conductivity, it is well to start with consideration of the conductance of a cathode in the absence of current in order to have a reference base.

The conductivity is

$$\sigma = nq\mu \tag{20}$$

To obtain the conductance per unit area, it is merely necessary to divide by the thickness of the cathode d, thus

$$g = \frac{nq\mu}{d} \tag{86}$$

The thickness d is about 10^{-2} centimeter, and μ has been measured by Pell to be about 10. It remains to determine n from Eq. (31), using DeVore's value for ϵ_D, i.e., $\epsilon_D = 1.4$ electron volts. Then, with $N_D = 10^{18}$, as before,

$$n = \sqrt{N_c N_D}\, \exp(\epsilon_D/2kT)\ 10^{15}\ cm^{-3} \tag{87}$$

and

$$g \sim 10^{-1}\ \text{mho}\ cm^{-2} \tag{88}$$

In other words, the resistance of a unit area of the cathode with no current is about 10 ohms. When current is drawn this resistance rises very markedly, and at high currents the heating due to cathode resistance becomes very large.

In practical cathodes, this problem is ameliorated by making cathodes porous.[21, 22] Part of the conduction then occurs through the pores where the mean-free-path of electrons is large. To illustrate the order of magnitude of pore conductivity:

Eq. (34) gives the density of the electrons in the pores

$$n_v = n \exp(-qX/kT) \tag{34}$$

With $n \sim 10^{15}\ cm^{-3}$, as above, and X = 0.5 ev,

$$n_v \sim 3 \times 10^{12} \tag{89}$$

The mobility is

$$\mu = \frac{8}{\pi}\,\frac{qL}{m\bar{v}} \tag{19}$$

With $l \sim 10^{-3}$ cm

$$\mu \sim 10^{4} \tag{89}$$

Then, $g \sim 5 \times 10^{-1}$ mho cm^{-2}, i.e., the pores have a conductivity about 5 times as large as that of a solid coating.

In cathodes made by coating silicon-bearing nickel, there is another resistance to be concerned about, namely interface resistance.[23, 24] Interface resistance arises from the diffusion of silicon out of the nickel and the subsequent reaction with BaO to form barium orthosilicate ($2BaO \cdot SiO_2$), which is an insulator. Depending on the thickness of the layer and on its activation (it is an impurity semiconductor), the resistance of the layer may range from a fraction of an ohm to thousands of ohms. The presence of interface resistance is easily detected because, in its presence, the current displays a decay with a time constant of about 1 microsecond. Because the decay of current due to donor depletion has a time constant of the order of a millisecond, the separation of interface resistance from bulk resistance is easy. When both are present, a diode has an equivalent circuit shown in Fig. 20. This circuit is valid at a given average current level; if the current level is changed, the circuit constants must also be changed. R_1 represents the initial sum of interface resistance and bulk resistance. C_2 with R_1 gives a time constant of about a microsecond and, thus, accounts for the interface decay. R_1 is usually about 5 times R_0 in the milliampere plate-current range. R_1 and R_0, together with C_1, have a time constant of about 1 millisecond, which corresponds to depletion decay. The latter decay is very embarrassing in dc oscilloscopes, because at an average current of 25 ma cm^{-2} the decay may amount to 5 per cent.

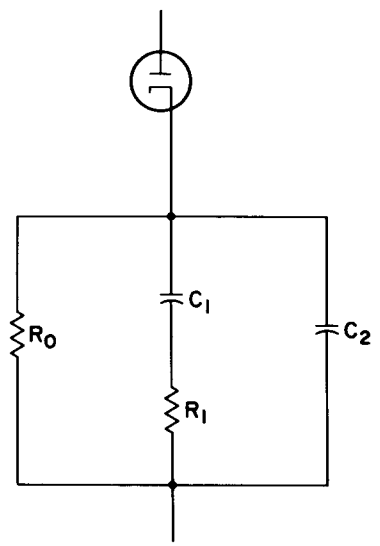


Figure 20. The Small-Signal Equivalent Circuit of a Diode with a Given Direct Current

The effect of cathode resistance on the current-voltage characteristic of a diode is illustrated in Fig. 21, which was drawn from data of Eisenstein.[25] To obtain these curves, Eisenstein measured the velocity distribution of electrons emerging from an aperture in the anode of a diode having an oxide cathode. The retarding potential required to stop electrons is equal to the potential drop in vacuum through which the electrons have fallen. Hence, the applied voltage minus the stop-

ping voltage is the potential drop through the cathode. The lower curve is the measured diode characteristic. The upper solid curve is the diode characteristic corrected for the cathode voltage drop. The dotted curve is the computed perveance line. It appears that the current is, in fact, space-charge limited and that the diode characteristic is deformed by the cathode voltage drop, which is 275 volts at an applied voltage of about 1500 volts.

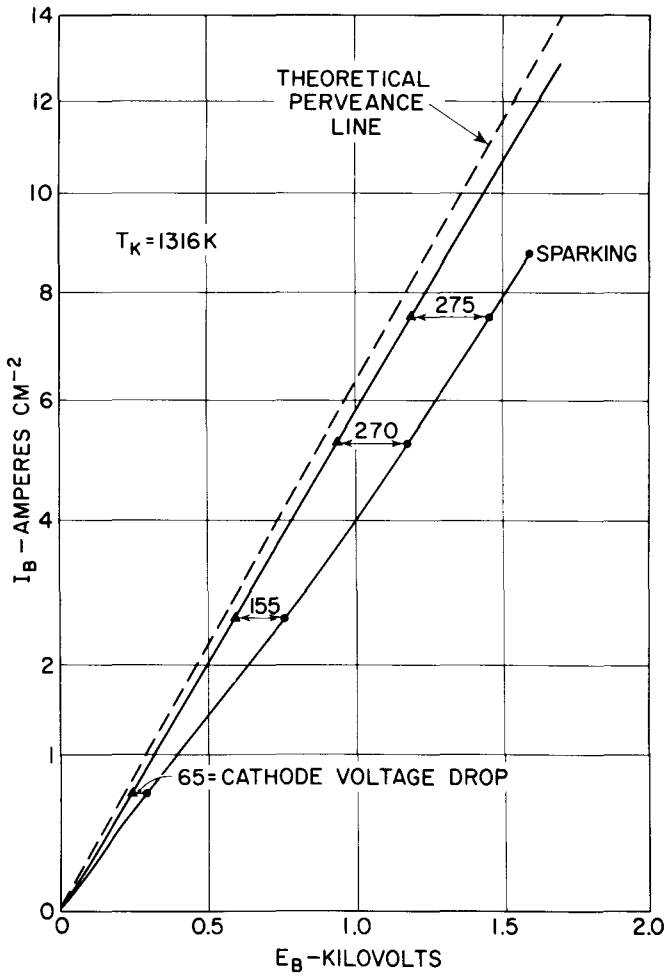


Figure 21. The Effect of Cathode Resistance on a Diode Characteristic (after Eisenstein)

The effect of cathode resistance on tube life is illustrated in Fig. 22. The data pertain to eight tubes and have been averaged for each point on the curves. The curve labeled I/I_0 is the ratio of emission to the emission at the beginning of life, both measured in a factory emission test set. The curve labeled P/P_0 is the ratio of power output into a standard load to the power output at the beginning of life. If it is supposed that the drops in emission and power output are due to an increase of cathode resistance, it is possible to compute the change in cathode resistance with life from the I/I_0 curve and then compute a P/P_0 curve for comparison with the experimental curve. The resistance change was computed from:

$$R_K = \frac{I_0^{2/3} - I^{2/3}}{\alpha^{2/3} I} \qquad (90)$$

where R_K is the cathode resistance and α is the perveance of the tube

Because the cathode resistance acts as a negative feedback resistor, it reduces the apparent transconductance g_m according to the relation

$$\frac{g_m}{g_{mo}} = \frac{1}{1 + g_{mo} R_K} \qquad (91)$$

where g_{mo} is the true transconductance of the triode

Finally, the ratio of power output to initial power output is

$$\frac{P}{P_0} \sim \left(\frac{g_m}{g_{mo}}\right)^2 \qquad (92)$$

When this computation was carried out, the computed values of P/P_0 were in agreement with the measured values. At the end of life, the cathode resistances of the tubes were measured and found to agree with the resistance computed from I/I_0. Finally, an attempt to measure emission was made. When the cathode resistance was corrected for, there was no evidence of emission limitation up to an ampere per cm^2. It appears that these tubes failed solely because of an increase of cathode resistance with life. It also appears that the factory test set did not measure actual emission; it measured a current limited by cathode resistance.

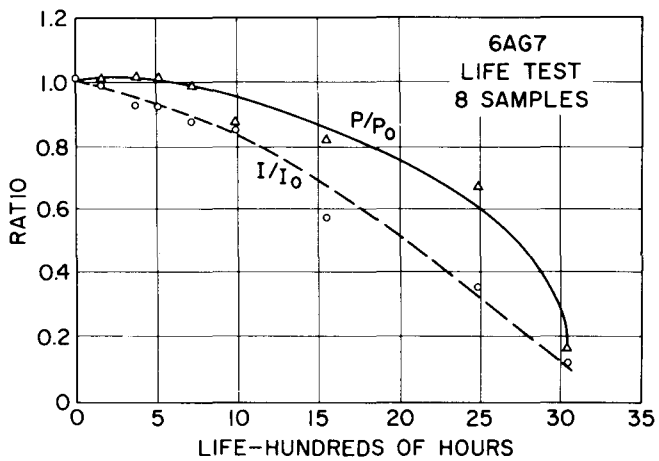


Figure 22. The Effect of Cathode Resistance on Tube Life

Pulse Operation of Oxide Cathodes

The work function of an oxide cathode is a function of the average current through the cathode. This relationship is illustrated in Fig. 23, which shows a set of Schottky plots obtained by operating the cathode at the average plate currents designated by I_{DC} in the figure and then delineating the Schottky curves with a short pulse. The zero-voltage intercepts of these curves are the field-free emissions, and are a measure of the work function. As the average current is increased from 3.5×10^{-6} to 13.2×10^{-3} amperes, the work function increases by about 0.11 electron volts.

Because the pulse operation of oxide cathodes is of importance in radar application, it is worth examining

the pulse behavior in a little detail.[20] The complete decay of cathode current is shown in Fig. 24. Although this tube was capable of an initial pulse emission of 40 amperes cm^{-2}, a measurement at this current level could not be made because the anode would have melted; hence, a low initial current was used. Note that the time constant is of the order of a millisecond and decay is almost complete in one second.

That the resistance of the cathode changes during pulse decay is shown in Fig. 25. This series of curves was obtained by interrupting the anode voltage for a few microseconds during a 145-microsecond pulse and measuring the current voltage characteristic during reapplication of the voltage. Notice that the departures from linearity of these curves start at very low voltages. If decay were due to emission limitation only, the curves should coincide in the low-current region and then break off abruptly at saturation. The early departure, which increases with elapsed time, shows that the internal resistance of the cathode increases during decay. After the long pulse, the E-I characteristics

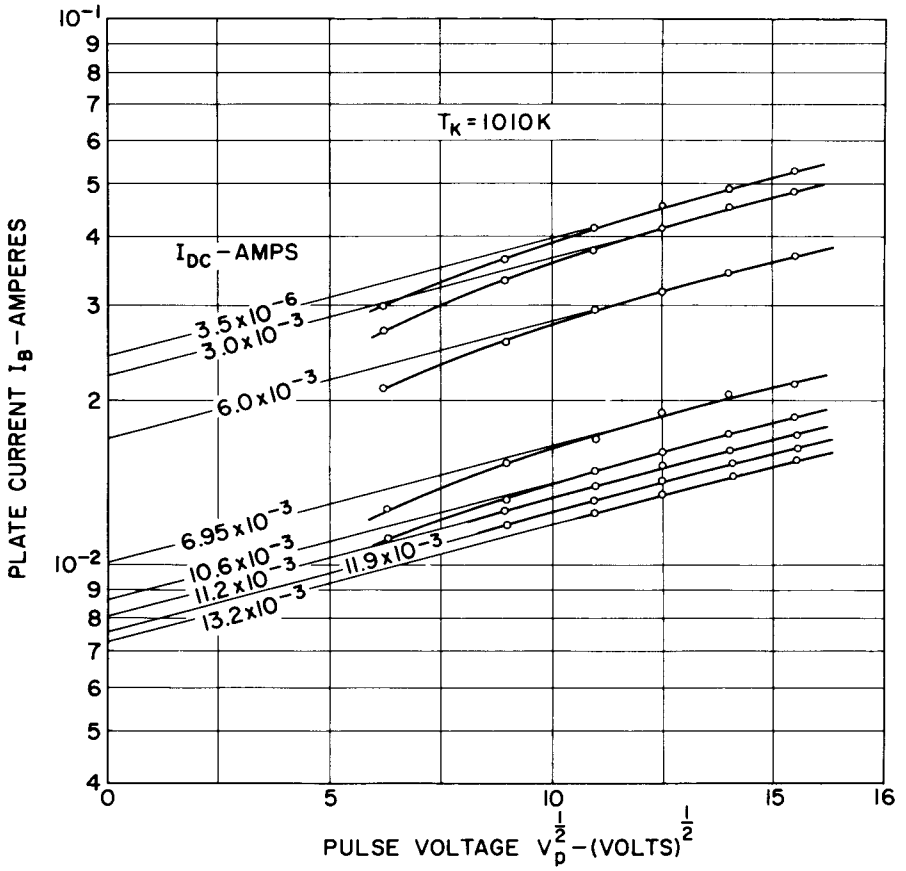


Figure 23. Schottky Curves Showing the Effect of Average Current on the Work Function of an Oxide Cathode

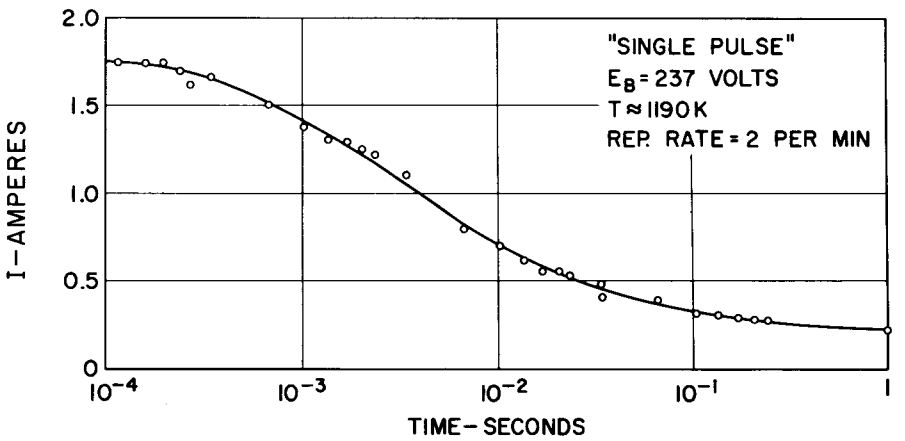


Figure 24. The Decay of the Current from an Oxide Cathode During a One-Second Pulse

were sampled with short pulses during recovery. Under these conditions, the curves of Fig. 25 are retraced in reverse order. Fig. 26 shows how the current decays and recovers during and after the 155-microsecond pulse. The variation of peak pulse current with duty is shown in Fig. 27. Notice that the peak pulse and dc emissions approach one another as the cathode temperature is reduced. This result is to be expected; as the temperature is lowered, so is the donor mobility, and at low-enough temperatures the donors will be "frozen in," and no pulse decay will occur.

Donors and Their Formation

The subsequent discussion relates to what the donors in the oxide cathode are, and how they are formed. The high diffusion rate of donors required to account for pulse decay suggested that they were F-centers— oxygen vacancies occupied by two electrons. To obtain data on this possibility, cathodes were activated at the receiving end of R. H. Plumlee's 60-degree mass spectrometer using the analyzer as a pure and controllable source of the reducing metals, Ba, Ca, Sr, and Mg.[26] Fig. 28 shows a typical result of depositing one of these metals (Ba) on an inactive cathode. When deposition starts, the electron current I rises markedly; when deposition stops, the electron current falls to an asymptotic value higher than the initial current. The rise during deposition is due to donors migrating through the depletion layer. The donors move slowly compared to the electrons, so that the electron current rises markedly in order to maintain space-charge neutrality. The effect is somewhat analogous to the gas amplification that occurs in a gas-filled photocell. The shape and magnitude of the deposition curves are independent of the metal used. Mg has an ionic radius of 0.78 A., Ba has an ionic radius of 1.43 A. Therefore, if the metals actually moved into the cathode, Mg should move very much more rapidly than Ba. That it does not suggests that neither is the donor, but that all of these reducing metals produce something else — perhaps F-centers -- that does move into the cathode. The increase in emission subsequent to deposition is proportional to the amount of metal deposited and is independent of

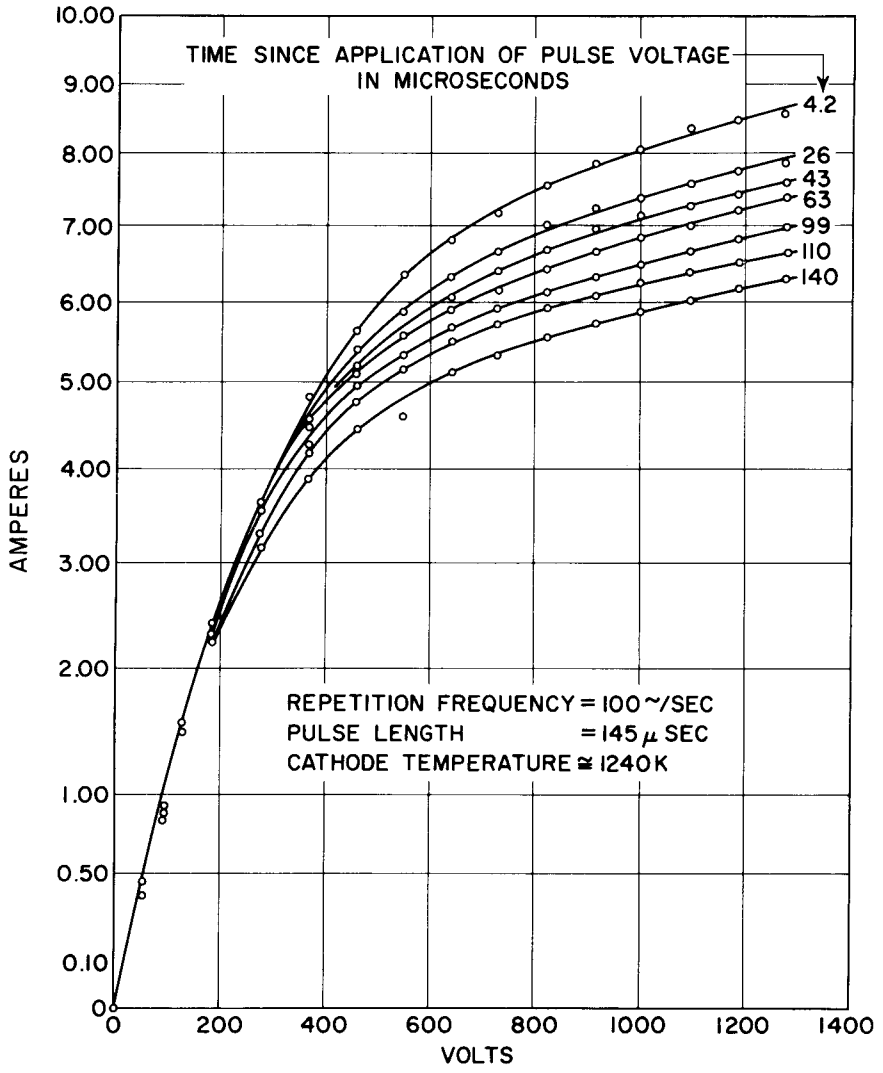


Figure 25. Current-Voltage Characteristics of a Diode as a Function of Elapsed Time During a 145- Microsecond Pulse

which metal is used. This result is shown in Fig. 29 in which $\Delta I/I_0$ is the fractional increase in emission and N is the number of monolayers of metal deposited. In computing N from the measured ion current and time of deposition, it was assumed that the actual cathode area was the projected area.

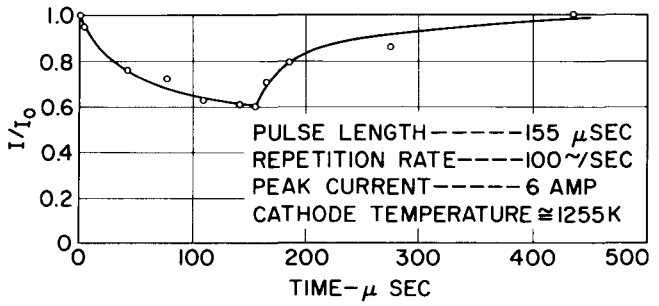


Figure 26. The Decay and Recovery of the Pulsed Emission of an Oxide Cathode (The available current during recovery was measured with a short sampling pulse)

As was noted in an earlier section, the amount of anything in a solid is related to the amount of the same thing in the gas phase by the relation

$$n_V = n_S \frac{N_C}{N} \exp(-\Delta H/kT) \qquad (10)$$

To study how the constituents of an oxide cathode are related to the gas-phase constituents of a vacuum tube, Plumlee installed a diode at the source end in his mass spectrometer.[27] The anode of the diode was apertured so he could examine the evaporation products of the cathode. The pressure in the spectrometer was less than 10^{-7} mm of Hg. (If this value seems like a good vacuum, recall that at this pressure about 3×10^{-2} monolayers of atoms are deposited on the cathode surface from the gas phase every second. If these atoms did not re-evaporate, they would increase the thickness of the cathode about 2 mils in 1000 hours.)

The equilibrium pressure of oxygen in contact with an oxide cathode proved of particular interest. It was found that the rate of evaporation of neutral 0_2 was proportional to the voltage applied to the diode. Becker had observed this behavior in 1929. Fig. 30 shows the results of Plumlee's measurements. The voltage-dependent evolution of 0_2 becomes less and less as the cathode activates (the state of activity is noted in the figure.) As was pointed out earlier in this section, the enhanced evaporation results from the electrolysis of oxygen ions to the surface. When oxygen evaporates, donors are formed. Because elimination of oxygen may seem tantamount to producing excess "free" barium, this experiment might be interpreted to show that the donors are barium atoms, a view widely held in the past. Recent experiments of Timmer at Cornell University show that this view is untenable.[28] Timmer has measured the Ba vapor pressure required to maintain excess Ba in BaO. His measurements show that in order to maintain the density of excess Ba in the cathode necessary to make the cathode active, a Ba vapor pressure of 10^{-6} mm of Hg is required. Hence, in any tube pumped to 10^{-7} mm of Hg or better, the donors cannot be free barium. A possible alternative will be discussed later.

Because the $C0_2$ peak observed in these measurements was larger than the 0_2, H_2, and H_20 peaks, the behavior of $C0_2$ was examined in somewhat more detail than the other peaks. Fig. 31 shows the behavior of the $C0_2$ peak as a function of time when the anode is turned on and off. When the anode voltage is turned on, $C0_2$ electrolyzes to the surface, enhances the evaporation rate, and the vapor pressure of $C0_2$ rises until the gas-phase thermal current of $C0_2$ to the surface just balances the

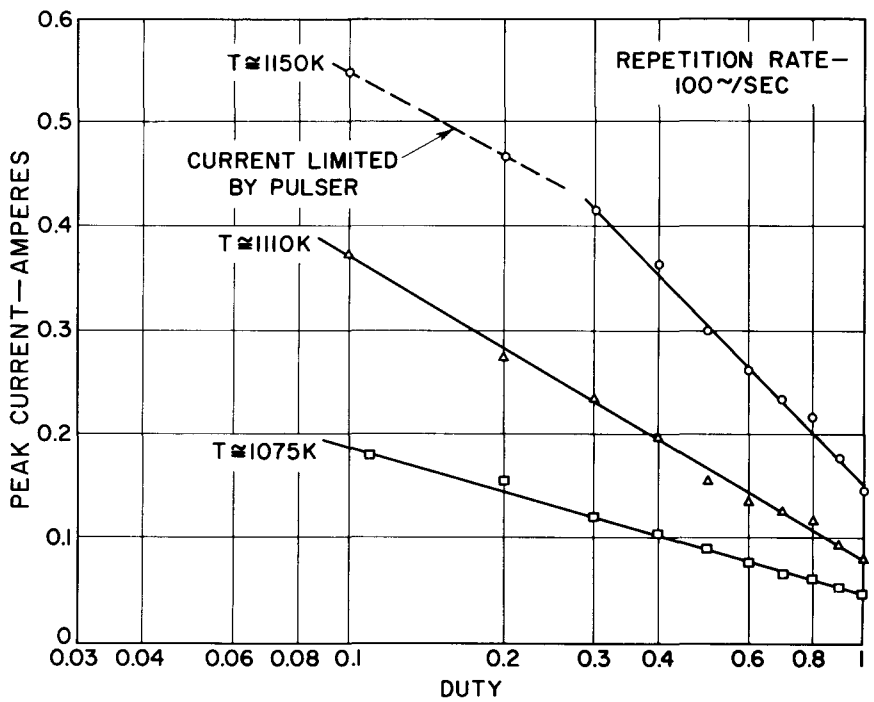


Figure 27. Peak Pulse Current vs. Duty for an Oxide Cathode

enhanced evaporation rate. Then, the peak height becomes constant. When the anode voltage is turned off, the evaporation rate drops and the cathode "soaks up" $C0_2$ until the initial situation is restored.

The H_2 (neutral) peak behaves in the same way. The water peak proved the exception – its behavior was the reverse of that of 0_2, H_2, and $C0_2$. This behavior is shown in Fig. 32. Aside from the spikes, which have yet to be explained, the general behavior shows that the cathode absorbs H_20 when anode voltage is applied. It was some time before an explanation of this behavior was found and a new class of donors had to be "invented" before the explanation would hang together. The "invention" and explanation is Plumlee's.[29,30] It can be summed up by the chemical equation

$$2\ Ba^{++}0^{--} + 2\ H_20 \rightarrow 2\ Ba^{++}(OH\ e)^{--} + 0_2 + H_2$$

Thus, water (in small amounts) reacts with BaO to form a donor with the concurrent release of 0_2 and H_2, and the behavior of 0_2, H_2, and H_20 is tied together. The hydrogen nucleus (a proton) carries a positive charge; therefore, it should electrolyze into the cathode on application of voltage in the forward direction of the diode. Furthermore, every oxygen ion in the lattice is a possible site for the proton. On this account alone the diffusion rate should be about 5 to 6 orders of magnitude faster than any vacancy diffusion mechanism. The activation energy for the diffusion of protons in ice has been measured and found to be about 0.25 electron volts. This agrees with the measurements of the "activation energy for pulse decay" (0.25 to 0.4 electron volts.) The diffusion rate of protons in BaO has been computed roughly, assuming that every oxygen ion is a possible site and that the activation energy is 0.25 electron volts. The result agrees with the diffusion rate required to account for pulse decay. Thus, Plumlee's hypothesis ties together a variety of phenomena qualitatively and, from present indication, quantitatively.

The oxide cathode is not yet completely understood, However, considerable progress has been made since World War II and the vast accumulation of knowledge of the oxide cathode in the fifty odd years since its discovery in 1904 is taking on a unity and coherence that

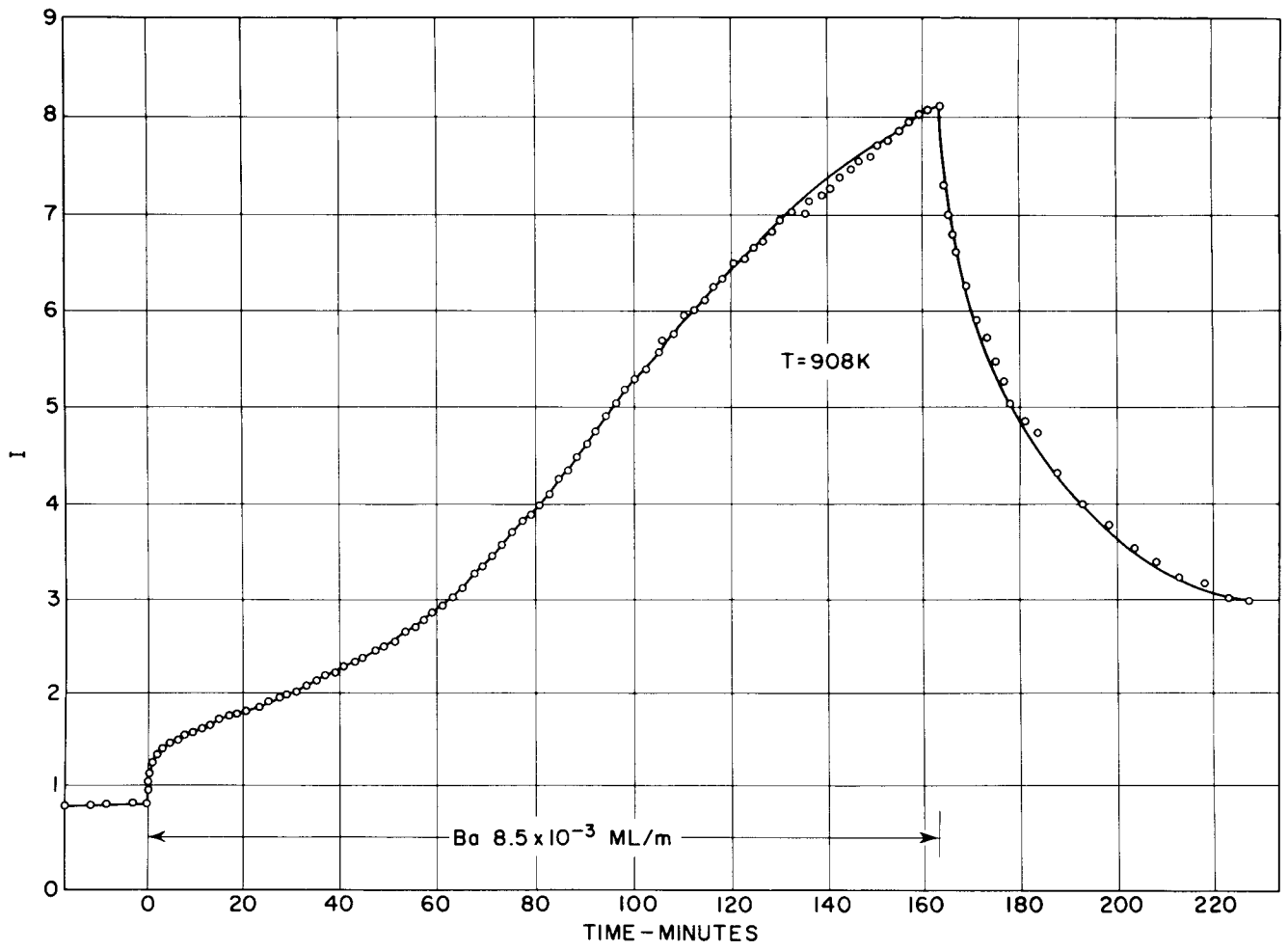


Figure 28. The Current Response of a Diode when Ba is Deposited on the Oxide Cathode. (The deposition rate is 8.5 x 10^{-3} monolayers per minute, computed on the assumption that the cathode is a smooth BaO crystal face.)

promises a complete understanding in the foreseeable future.

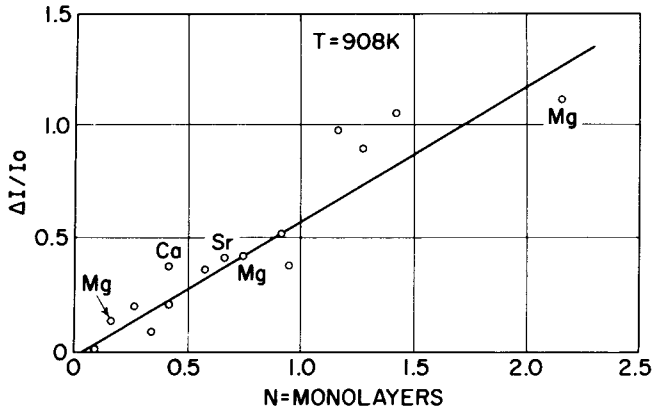


Figure 29. The Fractional Change in Activity of an Oxide Cathode as a Function of the Number of Monolayers of Reducing Metal Deposited (The unmarked points are for Ba)

APPENDIX A. THE "WAVE EQUATION" FOR ELECTRONS, AND THE DENSITY OF STATES

The Wave Equation

The simplest wave equation that can be written mathematically is the one-dimensional wave progressing in the x-direction.

$$\psi = \exp\left(-i\,\frac{2\pi x}{\lambda}\right) \tag{A1}$$

This function satisfies the differential equation

$$\frac{d^2\psi}{dx^2} + \left(\frac{2\pi^2}{\lambda}\right)\psi = 0 \tag{A2}$$

If λ is expressed in terms of De Broglie's relation

$$\lambda = \frac{h}{p} \tag{A3}$$

the propagation constant becomes

$$\frac{2\pi}{\lambda} = \frac{2\pi p}{h} = \frac{p}{\hbar} \tag{A4}$$

where $\hbar = h/2\pi$

Now express the momentum p in terms of the total energy ϵ and the potential energy U:

$$\frac{p^2}{2m} = \epsilon - U \tag{A5}$$

When the appropriate substitutions are made, the wave equation becomes

$$\frac{\partial^2\psi}{\partial x^2} + \frac{2m}{\hbar^2}(\epsilon - U)\psi = 0 \tag{A6}$$

This is the one-dimensional form of Schrodinger's wave equation. The above manipulation must not be regarded as a derivation of Schrodinger's equation. It is no more possible to derive this equation from first principles than it is to derive Newton's Laws from them. Schrodinger's equation should be regarded as an empirical formula, summing up the experience of the time, fortified by the vast number of predictions made from it which have been experimentally confirmed. Like Newton's laws, it has its weaknesses and fails to predict the relativistic effects now accounted for by its generalization, the Dirac wave equation.

The wave function ψ (x) is interpreted as a quantity such that ψ (x) ψ^*(x)dx is the probability that the electron will be found in the interval between x and x + dx (ψ^* is the complex conjugate of ψ).

A Simple Case of Quantization

Consider an electron in a one-dimensional potential

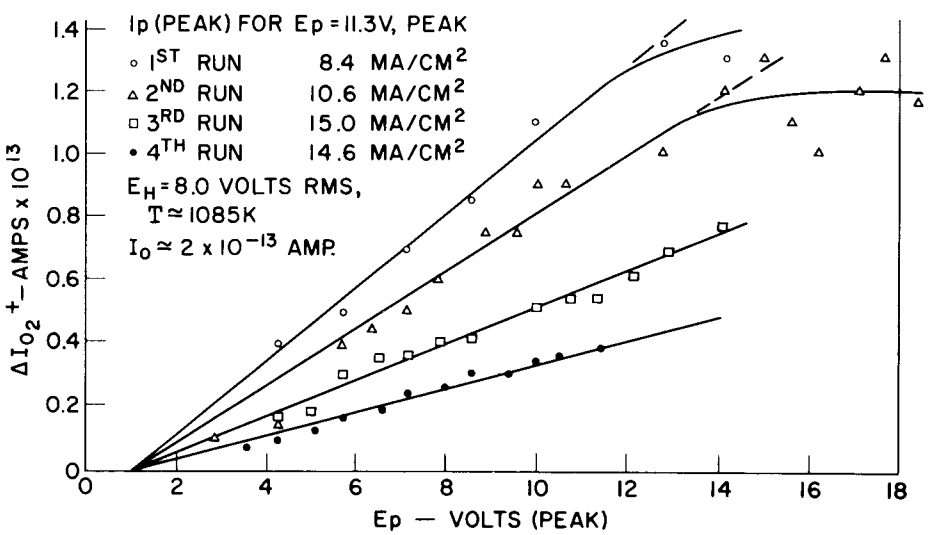

Ip (PEAK) FOR Ep = 11.3V, PEAK
1ST RUN 8.4 MA/CM2
2ND RUN 10.6 MA/CM2
3RD RUN 15.0 MA/CM2
4TH RUN 14.6 MA/CM2
EH = 8.0 VOLTS RMS,
T ≃ 1085K
Io ≃ 2 x 10-13 AMP.
ΔIO2 +- AMPS x 10^13
Ep — VOLTS (PEAK)

Figure 30. Oxygen Evolution from a BaO Cathode as a Function of Diode Voltage and Cathode Activity (The state of cathode activity for each run is noted in the table)

well such that U = 0 when $0 \leq x \leq a$, and U = ∞ outside of the well. Inside the well the wave equation is

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{2m}{\hbar^2} \epsilon \psi = 0 \qquad \text{(A7)}$$

And its solution is

$$\psi = A \cos \alpha x + B \sin \alpha x \qquad \text{(A8)}$$

Where

$$\alpha^2 = \frac{2m}{\hbar^2} \epsilon \qquad \text{(A9)}$$

Because the electron is inside the well and the wave function must be continuous at the boundary between inside and outside, ψ must vanish at the edges of the well, i.e., A = 0, $\alpha = \frac{n\pi}{a}$, n = 1, 2, 3, · · · .

Therefore,

$$\epsilon = \frac{\pi^2}{2m} \left(\frac{n\pi}{a}\right)^2 = \frac{1}{8m} \left(\frac{nh}{a}\right)^2 \qquad \text{(A10)}$$

and the possible energy states are discrete. The corresponding classical momenta are

$$p = \pm \sqrt{2m\epsilon} = \pm \frac{nh}{2a} \qquad \text{(A11)}$$

and are also discrete

The Density of States for a Free Electron

In the subsequent discussion of the distribution of

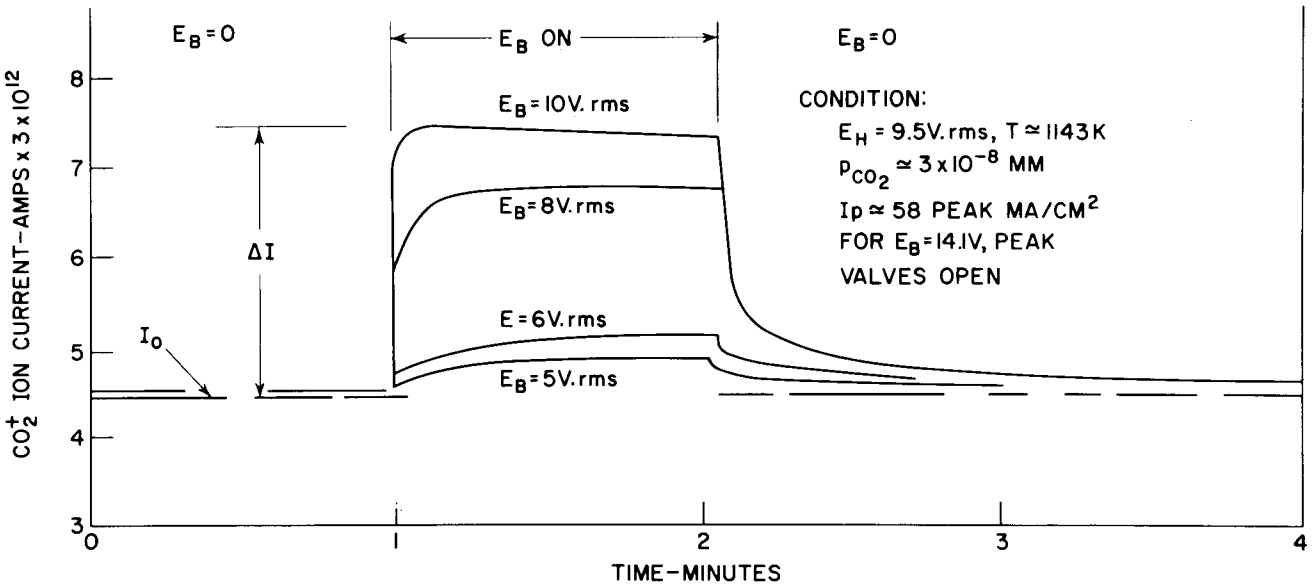


Figure 31. Carbon-Dioxide Evolution of a BaO Cathode as a Function of Time and Anode Voltage

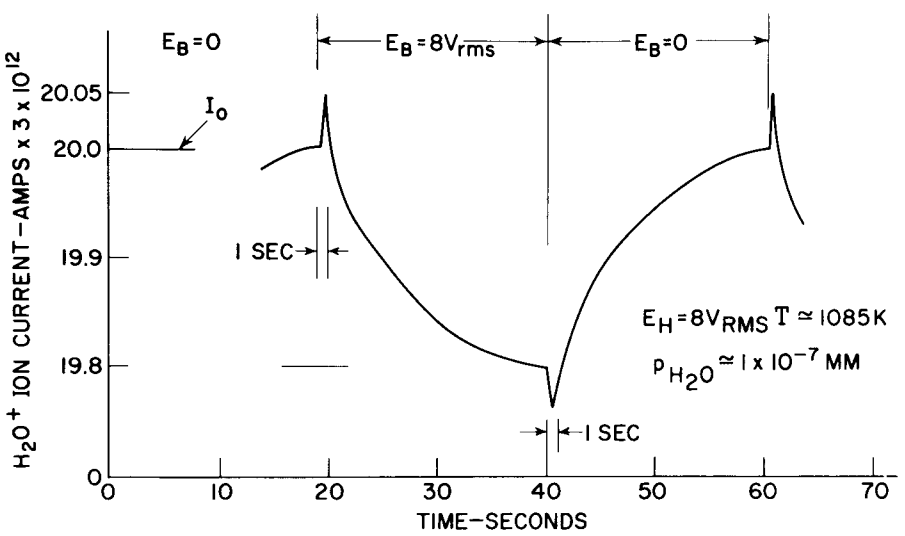


Figure 32. Water Evolution of a BaO Cathode as a Function of Time and Anode Voltage

"free" electrons among possible energy states, the density of states, i. e., the number of states in a given energy interval, must be known. This value may be obtained from the results of Eq. (A11) above.

If the size of the well—a in Eq. (A10)—is increased, the difference in energy between discrete states diminishes and ultimately becomes less than kT, the thermal energy of the electrons, and the electrons become "free." Then the number of states in the momentum interval Δp is

$$\Delta n = \frac{\Delta p}{h} a \qquad \text{(A12)}$$

The factor 1/2 in Eq. (A11) must be dropped in the enumeration because both plus and minus p correspond to the same energy. For the case of a three-dimensional box of dimensions abc, the relation corresponding to Eq. (A12) is seen to be

$$\Delta n = \frac{\Delta p_x \ \Delta p_y \ \Delta p_z \ abc}{h^3} \qquad \text{(A12a)}$$

where p_x, p_y, and p_z are the x, y, and z components of the momentum

Thus, the number of states per unit volume of space is the volume in momentum-phase space Δp_x, Δp_y, Δp_z divided by h^3. In other words, each state occupies a volume h^3 in phase space.

To compute the number of states in an energy interval, it is convenient to write Eq. (A12a) in spherical coordinates. The volume in a spherical shell of thickness dp and radius p is

$$dV_p = 4\pi \ p^2 \ dp \qquad \text{(A13)}$$

In terms of energy, this equation becomes

$$dV_p = 2\pi \ (2m)^{3/2} \ \epsilon^{1/2} \, d\epsilon \qquad \text{(A14)}$$

Then, the number of states in the energy interval between ϵ and $(\epsilon + d\epsilon)$ is

$$dN = \frac{2\pi \ (2m)^{3/2}}{h^3} \epsilon^{1/2} \ d\epsilon \qquad \text{(A15)}$$

which is the required relation.

APPENDIX B. FERMI-DIRAC STATISTICS

The Transitions Between Two "States"

Consider a group of N_1 states lying in an energy band not exceeding kT (k is Boltzmann's constant) in width centered at an energy ϵ_1 and a second similar band of N_2 states centered on an energy ϵ_2. Then the number of states occupied in the first band is the number of states multiplied by the probability of occupancy

$$n_1 = N_1 \left[\exp\left(\frac{\epsilon_1 - \epsilon_F}{kT} \right) + 1 \right]^{-1} \qquad \text{(B1)}$$

Similarly, the number of occupied states in the second band is

$$n_2 = N_2 \left[\exp\left(\frac{\epsilon_2 - \epsilon_F}{kT} \right) + 1 \right]^{-1} \qquad \text{(B2)}$$

Eliminating ϵ_F between the two equations, yields

$$n_1 (N_2 - n_2) \, e^{-\epsilon_2/kT} = n_2 (N_1 - n_1) \, e^{-1/kT} \qquad \text{(B3)}$$

By multiplying through by a constant, this expression may be written

$$n_1 (N_2 - n_2) \ W_{12} = n_2 (N_1 - n_1) \ W_{21} \qquad \text{(B4)}$$

This equation admits of a simple interpretation in terms of a bimolecular reaction: n_1 is the number of particles in the N_1 states, i. e., the number of particles available for transitions to the N_2 states; $(N_2 - n_2)$ is the number of unoccupied states to which transitions can be made; W_{12} is a rate-constant such that $N_1(N_2 - n_2) \ W_{12}$ is the rate at which particles make transitions from the N_1 states to the N_2 states. Then Eq. (B4) merely states that under equilibrium conditions there is no net exchange of particles between the states. The statistics yield no information on the magnitudes of W_{12} and W_{21} but state that their ratio is

$$\frac{W_{21}}{W_{12}} = \exp \left[-(\epsilon_1 - \epsilon_2)/kT \right] \qquad \text{(B5)}$$

The interpretation outlined above provides a simple and direct way of writing the equilibrium conditions for a system obeying Fermi-Dirac statistics.

Transitions to "Free" Energy States

Frequently it is necessary to consider transitions to the conduction band in a solid or transitions from a solid phase to a vapor phase. The electrons in a conduction band or the atoms or molecules in a vapor phase are essentially free, in that they may occupy states in a distribution of energy levels whose separation is small compared to kT. It is convenient to consider that all of the free electrons have one energy. The question then is what value to assign N_2 in Eq. (B4) when N_2 represents the number of "free" states. The value may be obtained as follows.

Eq. (A15) gives the number of levels available in the energy interval $d\epsilon$ at energy ϵ.

$$dN = \frac{2\pi \ (2m)^{3/2}}{h^3} \epsilon^{1/2} \, d\epsilon \qquad \text{(B6)}$$

The probability of occupancy is given by Eq. (B1). Hence, the number of particles in the interval $d\epsilon$ is

$$dn = 2\pi \frac{(2m)^{3/2}}{h^3} \frac{\epsilon^{1/2} \, d\epsilon}{\exp \frac{(\epsilon - \epsilon_F)}{kT} + 1} \qquad \text{(B7)}$$

The total number of particles is obtained by integrating ϵ from 0 to ∞. The integration is simple if $e^{\left[(\epsilon - \epsilon_F)/kT\right]} \gg 1$, which is true in most problems of concern. It should be noted that the energy is being meas-

ured from the lowest level among the free states and that ϵ_F must be negative for this inequality to hold. When the inequality does hold, Eq. (B2) may be approximated by

$$dn = 2\pi \frac{(2m)^{3/2}}{h^3} \exp \frac{(\epsilon - \epsilon_F)}{kT} \epsilon^{1/2} d\epsilon$$

Integration yields

$$n = N_c \exp\left(\frac{\epsilon_F}{kT}\right) \quad \text{(B8)}$$

where

$$N_c = \left(\frac{2\pi m\ kT}{h^2}\right)^{3/2} \quad \text{(B9)}$$

Thus, for transitions from states N_1, lying at an energy $|\epsilon_1|$ below the free states, to the free states, Eq. (B4) may be written

$$n_1 (N_c - n_1)\ W_{12} = n_1 (N_1 - n_1)\ W_{21} \quad \text{(B10)}$$

When $n_1 \ll N_c$, then $n_1 N_c\ W_{12} = n_1 (N_1 - n_1)\ W_{21}$ where

$$\frac{W_{21}}{W_{12}} = \exp\left(-\frac{|\epsilon_1|}{kT}\right) \quad \text{(B11)}$$

Studies of the magnetic properties of electrons reveal that they have a magnetic moment as though their charge were spinning about an axis and that if two electrons have their spins oppositely directed, both may occupy one energy state. Hence, for electrons

$$N_c = 2\left(\frac{2\pi m\ kT}{h^2}\right)^{3/2} \quad \text{(B12)}$$

LIST OF SYMBOLS

- A = Richardson constant
- a = a distance
- a = a constant = $n(\bar{F}l)^2/2qV$
- B = a constant
- b = a distance
- C = a capacitance
- c = a distance
- c = the velocity of light
- D = diffusion constant
- d = thickness
- E = electric field strength
- e = base of Naperian logarithims
- exp = exponential function
- F = force
- $f(x)$ = function of x
- g = conductance per unit area
- g_m = "apparent" transconductance (mutual conductance)
- g_{mo} = "true" transconductance
- ΔH = heat of sublimation
- h = Planck's constant
- i/h = $h/2\pi$ (also written $\hbar$)
- I = current
- I_o = initial current
- i = $\sqrt{-1}$
- i_s = secondary - emission current
- i_r = primary current
- j = current density
- j_D = diffusion current density
- j_e = electron-emission current density
- j_o = field-free-emission current density
- k = Boltzmann's constant
- L = mean free path
- l = a length
- m = particle mass
- m^* = effective mass
- N = Density of states; density of sites available to a constituent
- N_c = "density of states" in the conduction band of a solid or in the vapor phase
- N_D = donor density
- n = particle density
- n_o = particle density in a reference state
- n_s = particle density in a solid
- n_v = particle density in a vapor
- P = power output
- P_o = initial power output
- p = momentum
- q = electronic charge
- R_k = cathode resistance
- s = distance
- T = absolute temperature
- t = time
- U = potential energy
- V = voltage; energy in electron volts
- v = velocity; drift velocity
- $\bar{v}$ = mean thermal velocity
- $\bar{v}_n$ = mean thermal velocity normal to a surface
- W = work
- W_{mn} = transition probability from state m to state n per unit time
- x = distance
- α = perveance; attenuation constant
- β = a constant
- γ = density relative to density of water
- δ = secondary emission ratio
- ϵ = energy
- ϵ_D = activation energy for self-diffusion; energy gap between donors and conduction band
- ϵ_F = Fermi energy
- ϵ_g = band-gap energy
- η = depth of Fermi sea
- λ = wave length
- μ = mobility
- ν = frequency
- ν_o = threshold frequency for photoemission
- ξ = a displacement from equilibrium position
- ρ = charge density
- $\sim$ = approximately equal to
- σ = conductivity
- τ = mean collision time
- ϕ = work function
- χ = electron affinity
- ψ = a wave function
- ψ^* = its complex conjugate
- ω = frequency x 2π
- ω_p = plasma frequency x 2π

REFERENCES

1. Compton, K. T., and I. Langmuir, "Electrical Discharges in Gases," Rev. Mod. Phys., Vol. 2, p. 123, April 1930
2. PB 119240, Library of Congress, Washington 25, D. C.
3. Seifert, R. L. E., and T. E. Phipps, Phys. Rev., Vol. 53, p. 443, 1938; Vol. 56, p. 652, 1939
4. Munich, La Berge, and Coomes, Phys. Rev., Vol. 80, p. 887, 1950
5. Fowler, R. H., and L. Nordheim, Roy. Soc. Proc. A, Vol. 119, p. 173, 1928
6. Stern, T. E., B. S. Gossling, and R. H. Fowler, Roy. Soc. Proc., Vol. 124, p. 699, July 1, 1929
7. Hughes, A. L., and L. A. Dubridge, Photoelectric Phenomena, McGraw-Hill, New York, 1932
8. Zworykin, V. K., and E. G. Ramberg, Photoelectricity, Wiley and Sons, New York, 1949
9. De Vore, H. B., Phys. Rev., Vol. 83, p. 805, August 15, 1951
10. Kollath, R., Physik. Zeits., Vol. 38, p. 202, 1937
11. McKay, K. G., "Secondary Emission," a chapter in Advances in Electronics I, edited by L. Marton, Academic Press, New York, 1948
12. Ruthemann, G., Naturwissenschaften, Vol. 29, p. 648, 1941; Vol. 30, p. 145, 1942
13. Lang, W., Optik, Vol. 3, p. 233, 1948
14. Pines, D., and D. Bohm, Phys. Rev., Vol. 85, p. 338, 1952; Vol. 92, p. 608, 1953
15. Pines, D., Rev. Mod. Phys., Vol. 28, p. 184, 1956
16. Sternglass, E. J., Bulletin of Am. Phys. Soc., Series II, Vol. 1, No. 6, p. 801, June 21, 1956
17. Matheson, R. M., and L. S. Nergaard, Jour. Appl. Phys., Vol. 23, p. 869, 1952
18. Becker, J. A., and R. W. Sears, Phys. Rev., Vol. 38, p. 2193, 1931
19. Sproull, R. L., Phys. Rev., Vol. 67, p. 166, 1945
20. Nergaard, L. S., RCA Review, Vol. 13, p. 464, 1952
21. Loosjes, R., and H. J. Vink, Philips Res. Rpt., Vol. 4, p. 449, 1949
22. Forman, R., Phys. Rev., Vol. 96, p. 1479, 1954
23. Eisenstein, A., Jour. Appl. Phys., Vol. 20, p. 776, 1949
24. Nergaard, L. S., and R. M. Matheson, RCA Review, Vol. 15, p. 225, 1954
25. Eisenstein, A., Jour. Appl. Phys., Vol. 20, p. 776, 1949
26. Plumlee, R. H., R. M. Matheson, and L. S. Nergaard, RCA Review, Vol. 18, p. 385, 1957
27. Plumlee, R. H., RCA Review, Vol. 17, p. 190, 1956
28. Timmer, C., Jour. Appl. Phys., Vol. 28, p. 495, 1957
29. Plumlee, R. H., Jour. Appl. Phys., Vol. 27, p. 659, 1956
30. Plumlee, R. H., RCA Review, Vol. 17, p. 231, 1956

Oxide-Coated Emitters

C. P. Hadley

Lancaster

In 1903, Wehnelt[1, 2, 3] reported that the alkaline-earth oxides are efficient thermionic emitters of electrons. Over the fifty-odd years since this discovery, the oxide-coated cathode has been the subject of intensive investigation, and many workers have contributed to a better comprehension of its operation. Particularly rapid advances in understanding have occurred since the advent of the theory of semiconductors.[4] As a result, the phenomenon of thermionic emission from the oxide-coated cathode is now understood in principle but, in spite of extensive work, by no means in detail.

With regard to commercial use, the oxide-coated cathode is by far the leading thermionic emitter. In spite of its susceptibility to poisoning, ion bombardment, and damage by high voltages, its advantages of low cost, low temperature of operation, and relatively high current density have led to its use in an overwhelming majority of tubes requiring a copious supply of electrons. In recent years, several new emitters have appeared, such as the Philips L-cathode and the molded-nickel cathode. Although these types are finding uses in special applications, they have given no indication of monopolizing the field of thermionic emission. Indeed, no serious challenge has even been made to the supremacy of the oxide-coated cathode, nor is there any evidence that any successful challenge is imminent.

GENERAL DESCRIPTION

The typical oxide-coated cathode, shown in Fig. 1, consists of a solid solution of barium, strontium, and calcium oxides supported on a metallic base. The most useful base metals are various alloys of nickel. The cathode may be directly or indirectly heated. Typical operating temperatures range between 1000 K and 1100 K; 1075 K (brightness) is common in commercial cathode-ray tubes. The electron emission available under constant voltage conditions may be as high as 0.5 ampere per square centimeter, although 0.1 ampere per square centimeter is a more conservative value for long-life operation. Under microsecond pulse conditions, a current density of 10 to 50 amperes per square centimeter is not unusual. A typical oxide-coated cathode is made as follows: The coating material, composed of alkaline earth carbonates suspended in organic liquids in which a binder is dissolved, is applied to a nickel sleeve by spraying. After the cathode is mounted and placed in the envelope, several distinct processing steps are required for proper emission. These are: (1) evacuation of the envelope including baking; (2) burn-out of the binder; (3) decomposition of the carbonates to oxides (frequently called breakdown); and (4) activation (also known as aging).

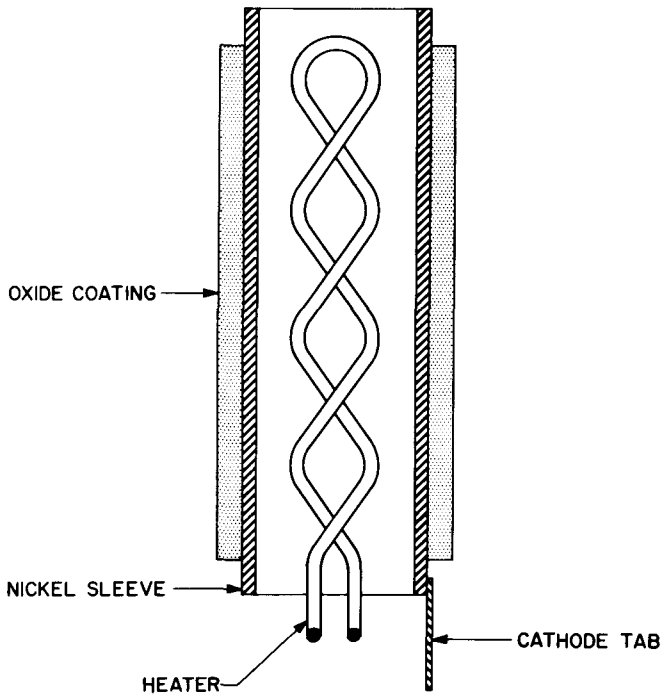


Figure 1. Cross Section of Indirectly Heated Oxide-Coated Cathode

THE CATHODE COATING

The carbonates for the cathode coating are usually precipitated from a water solution of the nitrates. Typical reactions are:

$$Ba(NO_3)_2 + Na_2CO_3 \rightarrow BaCO_3 + 2NaNO_3$$

$$Ba(NO_3)_2 + (NH_4)_2CO_3 \rightarrow BaCO_3 + 2NH_4NO_3$$

The first reaction produces a sodium-precipitated carbonate which is used in the majority of sprays. It has been often said, although not fully substantiated, that residual sodium which remains after washing of the carbonate is helpful in cathode activation. However, recent practice tends toward the use of the ammonium-

precipitated carbonates, which can be more readily cleaned of residual impurities.

A typical carbonate (RCA C-175) contains coprecipitated barium, strontium, and calcium carbonates in the molecular percentages of 49, 44, and 7 per cent, respectively. A typical cathode spray (RCA C-185) contains the same carbonate mixture and a binder of nitrocellulose, alcohol, diatol, and diethyl oxalate. In the preparation of the spray, the carbonate is thoroughly dried and then ball-milled with the binder.

The cathode coating is applied to the nickel sleeve by spraying. The coating may be a few thousandths of an inch thick and have a weight of 6 to 8 milligrams for each square centimeter of cathode area.

Various cathode sprays are used depending on the intended applications. The molecular ratio of the carbonates may be changed, for example, to obtain various emission levels. This effect was first shown by Benjamin and Rooksby[14]. Their data, repeated in Fig. 2, show that maximum initial emission is usually obtained with low percentages (20 to 30 per cent) of BaO. More recent results by C.H. Meltzer, show a much broader region of maximum initial emission. (See his chapter entitled, "The Oxide Coated Cathode.") Other workers have further increased initial emission by adding 5 per cent of CaO to the coating. Factors other than initial emission are involved, however, and the proportions used in any cathode spray are a compromise. Operating life, for example, is prolonged by an increase in barium content; whereas grid emission is decreased with an increase in calcium content; the cathode spray (C-185) described above, represents a suitable compromise for general use. Examples of cathode sprays for various applications are given below.

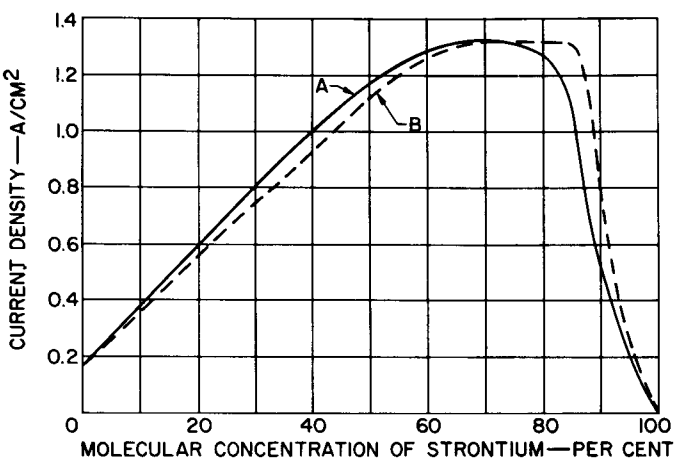


Figure 2. Dependence of Emission on Proportions of Barium Oxide and Strontium Oxide. Curve A: carbonate content before heat treatment; curve B, oxide content after activation[14]

1. For general use, when emission must be obtained readily and applied voltages are not too high, a spray which produces a fluffy coating (fast-drying) is desired; C-185, described above, is suitable.
2. The cathodes of tubes operating at high plate voltages require a smooth, dense coating. Such a coating is achieved by use of slow-drying solvents, such as amyl acetate, and carbonates having a small particle size. This last requirement can be satisfied with a mixture of singly precipitated barium carbonate and strontium carbonate.
3. If the structure of the tube is such that the grid runs hot, thermionic emission from the grid may be a problem. This grid emission may be decreased by decreasing the barium evaporation from the cathode, a result which is achieved by increasing the calcium content of the spray. Molecular concentrations of $CaCO_3$ as high as 20 per cent are occasionally used.

THE BASE MATERIAL

The functions of the metallic base are to support the cathode coating, to conduct heat and current to the coating, and to provide reducing agents for activation of the cathode. The last function, that of activation, will be explained in detail later. The essential point to be made here is that adequate emission is obtained only by virtue of activating centers within the coating, and that these centers are produced by a chemical reduction of the alkaline-earth oxides. The needed reducing agents may migrate to the cathode from remote parts of the tube, but the heat source of reducing materials is the metal sleeve which itself contains small quantities of impurities as alloys. The most common of the base metals is nickel alloyed with small amounts of silicon, magnesium, or carbon. The various cathode nickels have been arbitrarily classified as active, normal, or passive, depending on the quantity and character of the reducing agents present. The analysis of three typical alloys is given in Table I.

Table I

Material	Cathode Alloys (all values in per cent)		
	Active—599	Normal—220	Passive—499
Nickel and cobalt	99.0 min.	99.2 min.	99.8 min.
Silicon	0.15—0.25	0.01—0.05	0.01 max.
Magnesium	Present	0.01—0.1	0.005 max.
Carbon	0.04 max.	0.04 max.	0.05 max.
Copper	0.02 max.	0.2 max.	0.1 max.
Iron	0.02 max.	0.2 max.	0.1 max.
Manganese	0.2 max.	0.2 max.	0.002 max.
Sulfur	0.008 max.	0.008 max.	0.005 max.
Titanium	- -	- -	0.005 max.

The advantage of the active alloy is the ease with which emission is obtained, even under poor conditions of vacuum and cleanliness. There are, however, disadvantages which make the use of such an alloy undesirable except where necessary: (1) the cathode coating tends to peel from the base metal; (2) an interface material forms between the oxide and the base metal. When 599 alloy is used as the base, the interface material is barium orthosilicate (Ba_2SiO_4). The Ba_2SiO_4 interface is a poor electrical conductor, especially if current is not being drawn through it continuously, and, therefore, acts as a degenerative resistance in series with the cathode. The interface resistance is particularly undesirable in tubes designed for computer applications,

because such tubes may be cut off for a large proportion of the operating time.

Normal and passive alloys do not usually show interface resistance, and have less tendency to peel than the active alloys. Normal and passive cathodes are somewhat difficult to activate and should not be used when vacuum conditions are poor. Improved techniques of tube manufacture, however, make the use of active alloys unnecessary; modern tendency is toward the use of the purer nickels as base materials.

DECOMPOSITION OF BINDER AND CARBONATES

A vacuum bake is included early in the processing of a tube containing an oxide-coated cathode. If the bake is at more than about 275 C, the nitrocellulose binder is burned out. Following the vacuum bake, the coating is usually somewhat yellow or brown in color because of carbon left behind from the binder. If a Lucite binder is used, the binder sublimes without residue.

After the bake, heating current is applied to the cathode to decompose the carbonates. A typical reaction is:

$$BaCO_3 \longrightarrow BaO + CO_2$$

Some of the carbon dioxide reacts with the residual carbon from the binder to form carbon monoxide, and the cathode coating turns pure white. Both the carbon dioxide and carbon monoxide are pumped from the tube. The rate at which the breakdown occurs depends on the cathode temperature. Fig. 3 gives curves[5] of the vapor pressure of carbon dioxide during the decomposition as a function of the temperature for $CaCO_3$, $SrCO_3$, and $BaCO_3$. The barium carbonate is the most stable of the three compounds, and requires a temperature in excess of 1200 K for breakdown to occur in a reasonable time. A typical breakdown temperature would be 1250 K, and, for a small tube, the breakdown would be complete in 30 to 120 seconds. The breakdown is usually monitored with an ionization gauge; the completion of the reaction is indicated by a rapid fall in pressure.

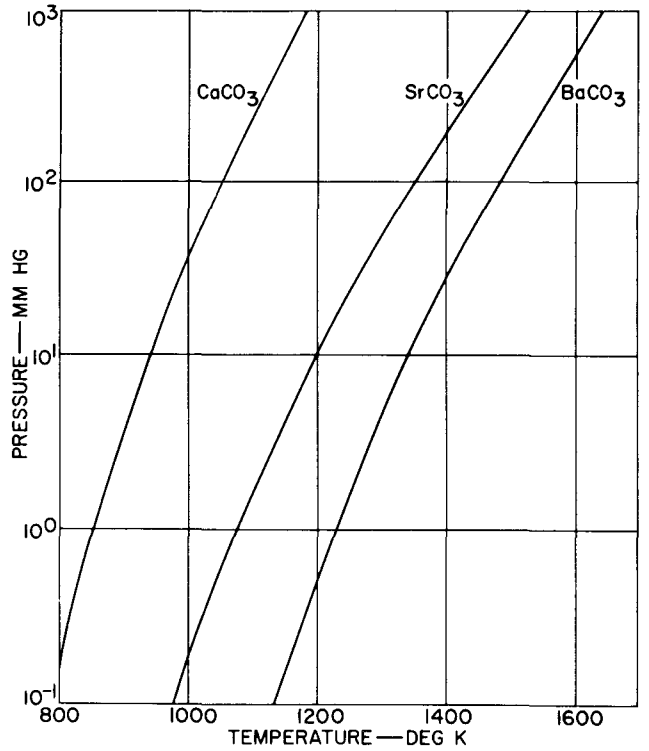


Figure 3. Vapor Pressure of Carbon Dioxide During the Breakdown Reaction

ACTIVATION

After the breakdown is complete, the cathode is activated by further heat treatment at temperatures near 1250 K. As explained previously, activation is essentially a chemical reduction of the cathode coating. There are three mechanisms by which such reduction may occur: (1) by dissociation of the oxide by ion bombardment, (2) by electrolysis of the oxide by passage of current, and (3) by chemical reduction by reducing agents. Chemical reduction is a very important mechanism and will be explained in some detail. The essential reaction involved is the reduction of barium oxide. One typical reaction is given below:

$$4BaO + Si \longrightarrow Ba_2SiO_4 + 2Ba$$

For a suitable reaction to occur, the reducing agent must satisfy the following requirements:

1. The reducing agent must be present in an available form. If the silicon or magnesium is in the cathode nickel as a compound of oxygen or sulfur (usually as inclusions at metal grain boundaries), for example, it will be bound too firmly to its anion to react with the barium oxide.
2. The reducing agent must be protected during the breakdown, otherwise it will react with carbon dioxide and be made unavailable. (If the reducer is alloyed with the nickel base, this requirement, as well as that of 1 above, is fulfilled.)
3. The reducer must diffuse through the base metal readily enough to be available for reaction.
4. The reduction must take place rapidly enough to be useful. Table II, devised by Rittner,[6] gives the equilibrium pressure of barium at 1000 K during the reaction between barium oxide and various elements. This equilibrium pressure is a measure of the rapidity with which the reaction progresses; i.e., the greater the equilibrium pressure, the more rapid the reaction. Although, of course, reducing agents other than those listed in Table II exist, they are either nearly inert, as far as the reduction of BaO is concerned, or are so active that the pressure of Ba produced is essentially that of Ba vapor in equilibrium with solid Ba.

As a rough rule of thumb, barium pressures no greater than 10^{-10} to 10^{-11} mm of mercury can be tolerated for cathodes of reasonable life. Thus, according to Table II, any cathode base composed entirely of a material more active than molybdenum would lead to excessive barium evaporation and a heavy interface.

If the reducing agent is included as an alloy of the base metal (for example, as silicon in nickel), the rate of reaction is reduced by two factors: (1) the presence of

the reducing agent as a compound (in the example, silicon would be present as Ni_3Si), and (2) the absence of a saturated solution of the reducer. Rittner[6] outlines the method of accounting for these factors and gives as an example the equilibrium pressure of barium over the reaction of BaO with 0.1 per cent (by weight) of silicon in nickel. At 1000 K, this equilibrium pressure is approximately 10^{-6} millimeters of mercury, a pressure which would be maintained only a short time during the initial operation of the cathode.

Table II

Reduction of Barium Oxide

Reducing Agent		Pressure of Barium (gas) mm. Hg
Thermal Disassociation		8×10^{-16}
Tin	(Sn)	4×10^{-15}
Iron	(Fe)	8×10^{-15}
Vanadium	(V)	9×10^{-13}
Zinc	(Zn)	2×10^{-12}
Gallium	(Ga)	4×10^{-12}
Molybdenum	(Mo)	6×10^{-12}
Chromium	(Cr)	4×10^{-10}
Tungsten	(W)	2×10^{-9}
Columbium	(Cb)	2×10^{-9}
Boron	(B)	1×10^{-8}
Manganese	(Mn)	1×10^{-8}
Tantalum	(Ta)	1×10^{-8}
Carbon	(C)	4×10^{-6}
Titanium	(Ti)	7×10^{-6}
Cesium	(Ce)	2×10^{-4}
Barium	(Ba(s) + Ba(g))	1×10^{-2}

MECHANISM OF OPERATION

Early workers on the oxide cathode recognized that high emission was associated with the presence of excess barium.[7] It was believed that the excess metal formed a dipole layer on the surface of the oxide and, thus, reduced the work function. More recent work has emphasized the semiconductor nature of the cathode, and has associated the principal activating centers with oxygen deficiencies.[8] Still more recently, Plumlee[9] has indicated that the activating centers may be more complex, possibly involving the OH^- radical. Experimental measurements of the energy of the various levels of the band structure of BaO have yielded a variety of results. Nergaard[10] has summarized them as follows (see Fig. 4):

1. Band gap: about 5 electron volts.
2. Electron affinity: about 0.5 electron volt.
3. Valence-band exciton states: about 1.2 electron volts below the conduction band.
4. Impurity levels: 3.5, 2.4, 2.0, 1.4, and 0.8 electron volts below the conduction band. (The levels at 1.4 and 2.4 electron volts are probably most certain.)

The two important electrical characteristics of the cathode are its thermionic emission and its conductivity. These characteristics may readily be calculated[11]. The

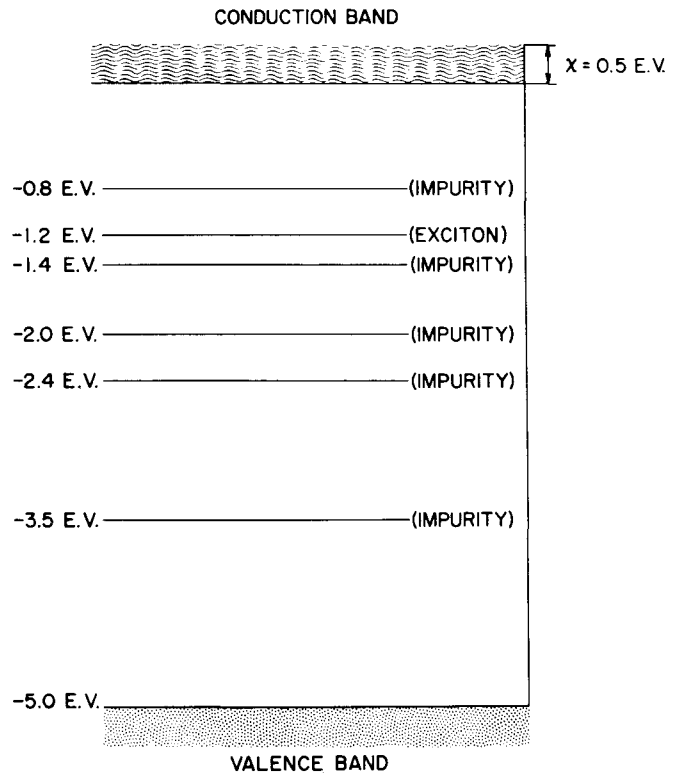


Figure 4. Energy-Level Structure of Oxide Cathode[10]

density of electrons in the conduction band n_c is given by

$$n_c = N_c e^{q\zeta/kT} \quad , \quad N_c = \frac{2}{h^3}(2\pi m k T)^{3/2} \tag{1}$$

Here N_c represents the density of available states in the conduction band, ζ is the energy of the Fermi level (measured from the bottom of the conduction band) q is the electron charge, k is Boltzmann's constant, T is the absolute temperature, h is Planck's constant, and m is the mass of the electron. The energy of the Fermi level ζ may be calculated from the expression

$$n_c + \sum_i \frac{N_i}{1 + e^{q(E_i - \zeta)/kT}} = \sum_i n_{oi}$$

where N_i is the density of states in the forbidden band at the energy E_i, and n_{oi} is the density of electrons in the i^{th} state at a temperature of absolute zero.
The conductivity of the coating σ is given by

$$\sigma = n_c q \mu \tag{2}$$

and thermionic emission current density is

$$j_o = n_c q \sqrt{\frac{kT}{2\pi m}} e^{-q\chi/kT} \tag{3}$$

In the above expressions, μ is the electron mobility, and χ is the electron affinity.

The problem in making any calculations is to determine the proper model to use. The experimentally de-

termined model of energy levels described above is complicated; and, in any event, the density of the various levels is not well known. A simplified model of barium oxide which explains many, although not all, of the experimental results, is as follows: As stated previously, the band-gap is about 5 electron volts and the electron affinity about 0.5 electron volts. It is assumed that only one activation level, a donor level, is important and that it is located at about 1.4 electron volts below the conduction band. In the notation being used, $\chi = 0.5$ electron volts, $E_i = E_b = -1.4$ electron-volts, and the density of the activation levels is given by $N_i = N_b$. With these assumptions, the emission and conductivity equations become:

$$j_o = \left[\frac{qk^{5/4}(8\pi m)^{1/4}}{h^{3/2}}\sqrt{N_b}\right] T^{5/4} e^{-q\phi_o/kT} \qquad (4)$$

$$\sigma = \left[\frac{\sqrt{2}}{h^{3/2}}(2\pi mk)^{3/4}\mu q\sqrt{N_b}\right] T^{3/4} e^{qE_b/2kT} \qquad (5)^*$$

ϕ_o, the work-function, is given by $\phi_o = \chi - E_b/2$

Similar equations containing slightly different band-structure values apply to a triple-oxide cathode. Typical constants for such a cathode are:

$\phi_o = 1$ electron volt
$N_b = 10^{23}$ levels/cubic meter
$j_o = 20$ amperes/square centimeter
$\sigma = 10^{-2}$ mhos/centimeter

The current density given above is well in excess of the emission usually drawn, and applies more correctly to pulsed emission. The reason that dc emission is lower than pulsed emission by at least two orders of magnitude is ascribed by Nergaard[8] to the mobility of the donors. According to this picture, donor levels which are ionized by thermal excitation of electrons to the conduction band experience a force away from the cathode surface when current is drawn. The result is that N_b of Eqs. (4) and (5) becomes quite small near the emitting surface, and the cathode conductivity becomes low in this region.

Another factor of importance in the operation of the cathode is the interface resistance. As described previously, cathode nickels which contain silicon lead to the formation of an interface compound. A sketch of such a situation is shown in Fig. 5. The interface material is itself a semi-conductor and exhibits a conductivity of the sort expressed by Eq. (5). A typical interface layer may be of the order of 10^{-3} centimeters in thickness and may, if well-activated, exhibit a conductivity at 1000 K of about 10^{-6} mhos/centimeter[12]. In some cases, the resistivity may be much higher. For example, the resistance of the interface tends to be higher if current is not drawn continuously from the cathode.

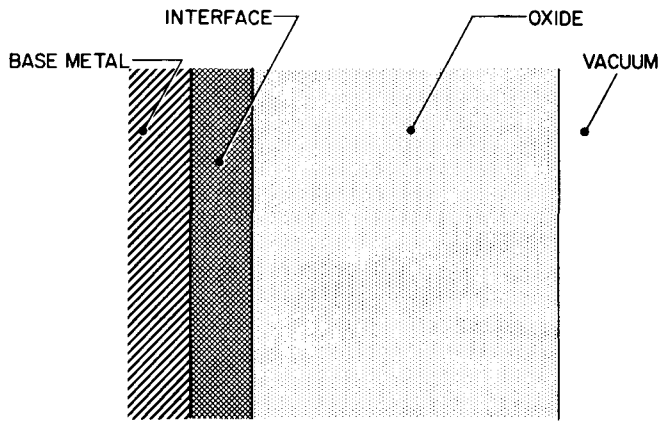


Figure 5. Diagram of Oxide Cathode with Interface[12]

THE MEASUREMENT OF THERMIONIC EMISSION

A measurement of thermionic emission must be made on each manufactured tube which contains an oxide-coated cathode; such measurement is by no means straightforward. Indeed, most so-called "emission" measurements are relative determinations, the results of which depend on many factors other than the condition of the cathode. Such factors include tube geometry and cleanliness of collecting electrodes.

The most practical method of making a true measurement of the thermionic emission of an oxide-coated cathode is to connect the tube involved as a diode. In the elementary theory usually given for the current-voltage characteristics of a diode, two regions are recognized. For the lower voltages, the current is said to be space-charge limited, and the current varies as the 3/2 power of the voltage. For higher voltages, the emission is said to be temperature limited and is essentially constant. On the basis of this simple theory, a satisfactory emission measurement can be made by the application of a collection voltage which is sufficiently high to reach the saturation region.

Unfortunately, a number of additional effects occur which are of importance and which complicate the measurement considerably. These additional effects, listed below, are described in detail later.

1. The initial velocities of the emitted electrons alter the simple space-charge theory.
2. The application of a high collection field lowers the surface barrier over which the electrons must pass to be emitted and thus eliminates a true temperature saturation.
3. The electrical resistance of the cathode coating causes a voltage drop during emission measurement. As a result, the actual voltage difference between the cathode surface and the anode is less than the applied voltage.
4. The cathode may be poisoned during testing. Cathode poisoning raises the cathode surface barrier and decreases emission.

*Under some circumstances the current may be carried as a flow of space charge through the pores of the cathode coating. In such cases, Eq. (5) does not apply.[15]

5. Because of the difference in work functions of the cathode and anode, a contact difference in potential acts in series with the tube.
6. Some types of cathodes (oxide-coated cathodes in particular) give a higher emission for microsecond pulses than they do for steady-stage voltage conditions.

In the previous section (Mechanism of Operation), the thermionic emission equation for the oxide cathode was derived. Any proper measurement of emission must be characteristic only of the cathode material and hence must be some form of evaluation of Eq. (4). The experimental measurement, therefore, must duplicate the assumptions under which the emission equation was derived. It is not profitable to list all of the assumptions here, because most of them are well-enough satisfied in ordinary procedure. Two of the assumptions, however, necessitate special precautions. These precautions are:

1. The cathode condition (N_b and ϕ_o) must remain constant during a measurement.
2. The electrons being measured must not escape over a barrier either greater or less than χ. This implies that the electric field at the surface of the cathode must be zero.

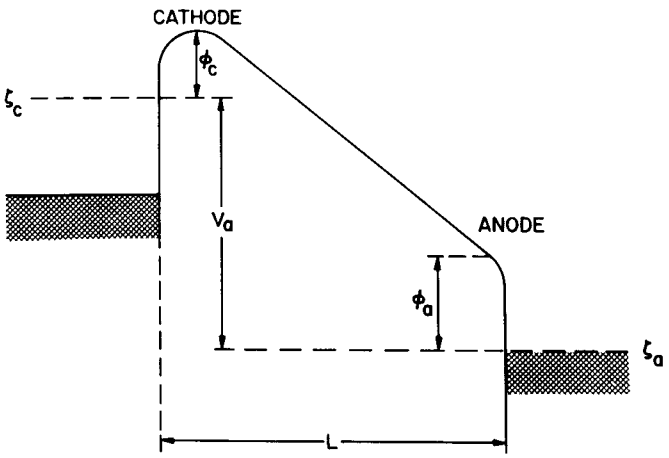


Figure 6. Potential Distribution in a Diode (no space charge)

Fig. 6 gives a potential-energy diagram for an electron in a plane-parallel diode, in which the anode is held positive with respect to the cathode by V_a volts. It is assumed that the electron-charge density in the cathode-anode space is negligible. The diagram shows the surface barriers, ϕ_a and ϕ_c, and the Fermi levels of the two electrodes. The electric field in the cathode-anode space is given by

$$\left[V_a - (\phi_a - \phi_c) \right] / L$$

where L is the cathode-anode spacing.

The difference between the anode and cathode work-functions, $\phi_a - \phi_c = V_c$, is called the contact difference in potential, and must be taken into account when the voltages applied to the diode are small. The actual potential difference across the vacuum gap for an applied voltage of V_a is $V_a - V_c$.

If the cathode coating has an appreciable resistance R_k, a potential difference $R_k i_k$ is formed across it by the cathode current i_k. The vacuum gap potential difference then becomes

$$V_a - V_c - R_k i_k \tag{6}$$

The electrons within the cathode-anode space of the plane-parallel diode must obey the usual laws of mechanics and electrostatics. Therefore,

$$\nabla \cdot \vec{D} = \rho \qquad \frac{mv^2}{2} = \frac{mv_o^2}{2} + q\psi$$

$$\vec{j} = nq\vec{v} \qquad \vec{D} = \epsilon_o \vec{E}$$

D is the electric displacement, ρ the charge density, v the velocity of an electron at the potential ψ, v_o the velocity of an electron at zero potential, $\vec{j}$ the current density, n the density of electrons in the vacuum space, ϵ_o the permittivity of free space, and $\vec{E}$ the electric field. If it is assumed that the initial electron velocity is zero, these equations may be combined to yield the following expression for the cathode current density, j_1:

$$j_1 = \frac{4\epsilon_o}{9} \sqrt{\frac{2q}{m}} \frac{V^{3/2}}{L^2} \tag{7}$$

Eq. (7) is the familiar space-charge relationship.

The above equation is, of course, somewhat idealized; a much more complete analysis is given by Nottingham.[11] A sketch of the actual potential distribution between the cathode and anode, neglecting V_c, is given in Fig. 7. The equations which describe the current density j_2 are as follows:

$$j_2 = j_o e^{-qV_s/kT} \tag{8}$$

$$j_2 = \frac{4\epsilon_o}{9} \sqrt{\frac{2q}{m}} \frac{(V + V_s)^{3/2}}{(L - l)} \tag{9}$$

V_s is the height of the potential barrier due to space charge (see Fig. 7), and l is the distance of the top of the barrier from the cathode surface.

If the applied voltage is varied, the potential distribution changes, as shown in Curves B and C of Fig. 7. As V is made more positive, the potential barrier becomes smaller and moves toward the cathode. At a critical voltage, the field at the cathode is zero, V_s and l are zero, and $j_2 = j_o$, which is the current described by Richardson's equation. One important method of determining emission, which will be described later, involves a technique for increasing V until $j_2 = j_o$.

In the introduction to this paragraph, an idealized theory was described as a simple explanation for the current-voltage curve of a diode. The subsequent discussion has shown some of the fallacies in the simple explanation. In reality, the characteristic curves look more like those shown in Fig. 8. Curve A shows the case when the cathode resistance is negligible, curve B when appreciable resistance is present. The theory of Eqs. (8) and (9) is meaningful when the cathode current is limited by space charge, that is, when the

applied voltage is rather low. Eq. (7) applies only when the field at the surface of the cathode is exactly zero, which occurs only at a single current value for any given diode. As the voltage applied to the diode is gradually raised, a semi-saturation of current is reached as shown in curve A of Fig. 8. In this region, called the Schottky region, additional theory is required for explanation.

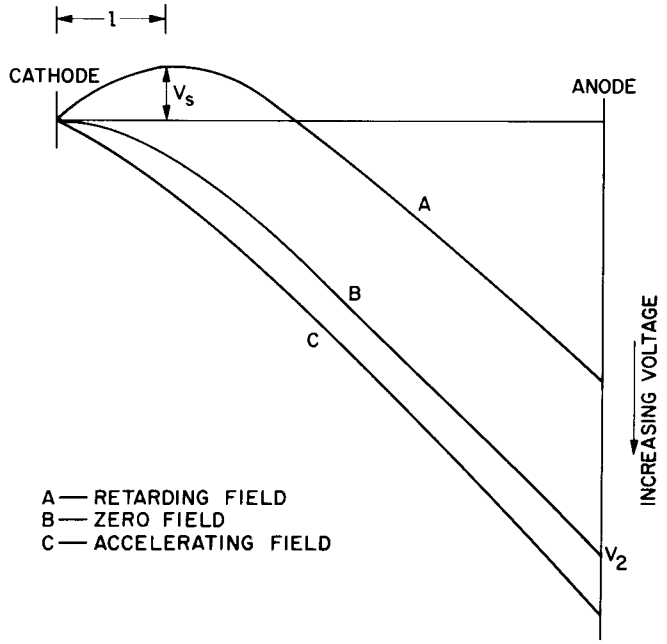


Figure 7. Potential Distribution in a Diode (space charge): curve A, retarding field; curve B, zero field; curve C, accelerating field

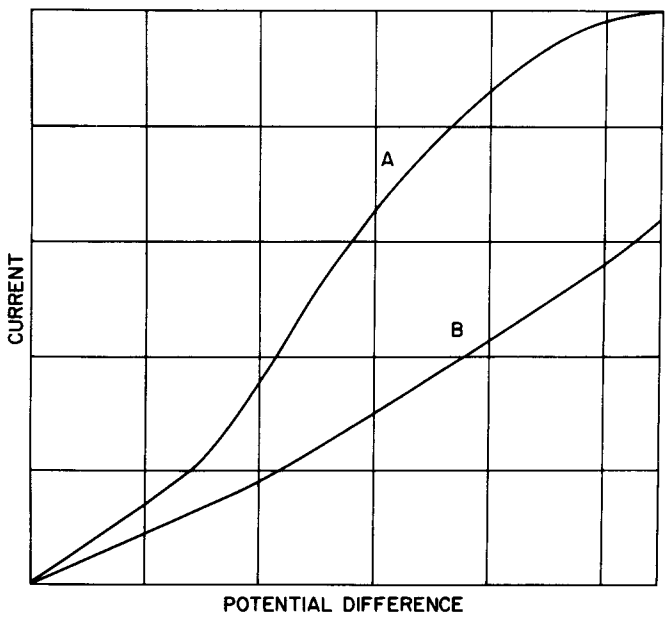


Figure 8. Current versus Voltage Characteristics for a Diode: curve A, zero cathode resistance; curve B, appreciable cathode resistance

As shown in Fig. 7, diode voltages beyond the space-charge region cause an accelerating field at the surface of the cathode. This accelerating field lowers the surface barrier (χ in Fig. 4), and allows electrons of lower-than-usual energy to be emitted. The calculation of the amount by which the barrier is lowered was first made by Schottky[13], who made the assumption that the force exerted by the emitter on an emitted electron is the familiar image force. The theory is, therefore, accurate only when the electron is at some distance from the emitter. The lowering of the work-function barrier is given by

$$\Delta\phi = \sqrt{\frac{qE}{4\pi\epsilon_0}}$$, where E is the field at the surface of the cathode

The current density for the accelerating condition is, therefore

$$j_3 = j_0 e^{q\Delta\phi/kT} \qquad (10)$$

For any diode, E is proportional to V; therefore, a plot of $\ln j_3$ versus $\sqrt{V}$ should give a straight line having an intercept $\ln j_0$. This method is sometimes used for the determination of the field-free current.

It is important that the condition of the cathode remain constant during measurement. Any poisoning will cause ϕ_0 to increase and N_b to decrease. Because j_0 is more sensitive to ϕ_0 than to the other variables, the net effect of the poisoning is a lowering of emission. There are three factors most important to cathode poisoning.

1. Even for well-processed tubes, excessive heating of the anode causes a tube to become gassy. The effect can be minimized if the anode power is kept to a minimum. The two requirements of low anode power and zero-field current tend to work in opposite directions. If zero-field current is drawn under dc conditions, it is usually necessary to reduce the cathode temperature. For temperatures nearer the operating range, the anode power may be kept low by the application of voltage in microsecond pulses.
2. Any cathode (and in particular an oxide-coated cathode) evaporates material to the anode during life. These evaporated materials, frequently oxides, can be decomposed by electrons having more than about 10 volts of energy. The decomposition products are able to poison the cathode.
3. Poisoning effects act much more strongly as the cathode temperature is lowered.

In addition to the effects discussed above, which apply to all cathodes, there are some special effects which are characteristic especially of oxide-coated cathodes. These effects result from the semiconductor nature of these emitters. In the section on thermionic emission theory, it was noted that the emission from semiconductors was due to the presence of donor levels. The donors are mobile at the temperature of operation, and are moved toward the base material by the electric field which is present during the drawing of current. Two of the effects of donor mobility which are applicable to

this discussion are:

1. Pulse emission for oxide cathodes is from 10 to 100 times as large as dc emission.
2. The Schottky theory of Eq. (10) may not hold. If emission data are plotted as $\ln j_3$ versus $\sqrt{V}$, the slope of the curve for a plane-parallel diode is as follows.

$$\frac{d}{dV^{1/2}} \ln j_3 = \frac{q}{kT} \left(\frac{q}{4\pi \epsilon_0 L} \right)^{1/2}$$

Similar equations are found for tube geometries other than plane-parallel. Metallic emitters and emitters such as thoriated tungsten perform as predicted by Eq. (10), but oxide-coated cathodes show an anomalously high slope (3 to 4 times that expected). This effect casts doubt on the validity of using a Schottky plot for the determination of field-free emission for oxide-coated cathodes.

On the basis of the above discussion, the most satisfactory simple way to properly measure the thermionic emission of an oxide-coated cathode is:

1. Apply voltages to the diode-connected tube. Draw currents within, and slightly above, the space-charge region.
2. Make a plot of the current to the 2/3 power versus the applied voltage. (Eq. 7).
3. Correct the plot for cathode-coating resistance. This correction may frequently be made by a determination of what coating resistance must be assumed in order that the space-charge current satisfies geometric considerations.
4. Find the field-free emission current at the point where the plot of the measured current departs from the space-charge line.

It is tedious to make the space-charge plot point-by-point, and it is impossible in production. The next best expedient is to make a single pulsed-emission measurement at a voltage sufficiently high that the current does not change too rapidly with voltage. The emission so measured will be several times higher than the field-free value.

The least satisfactory method of all is to measure current with a moderate voltage applied to the collector. The accuracy of such a measurement depends primarily on geometry and on the resistance of the cathode. Since Eqs. (4) and (5) show that both the emission and the resistance vary with cathode activity, the current will be some measure of the activity of the cathode if the geometry can be assumed constant from tube to tube.

DYNAMIC EQUILIBRIUM AND POISONING

Although the previous paragraphs have discussed separately the various mechansims occurring in oxide cathodes, the condition of the cathode at any time is determined by a dynamic equilibrium of all the mechanisms in progress. It is convenient to classify the mechanisms as follows:

Activating Mechanisms

1. Migration of reducing agents through base metal.
2. Reducing action between the oxide coating and various reducing agents present.
3. Diffusion of the activating centers throughout the oxide coating.
4. Reduction by electrolysis.

De-Activating Mechanisms

1. Evaporation of activating centers.
2. Oxidation of the coating by undesirable impurities such as oxygen, water vapor, and chlorides.
3. The evolution of gases from tube parts.
4. The reduction of compounds on electrode surfaces by electron bombardment.

Destruction of Barium, Strontium, or Calcium Oxides

1. Reaction with materials such as water vapor, chlorides, and carbon dioxide.
2. Destruction by ion bombardment.
3. Evaporation of the coating.

The cathode will become useless for emission if any of the activating mechanisms is interrupted unduly, or if the undesirable reactions are present to any extent. According to the discussion following Eq. (4), a well-activated cathode has an activator concentration of about 10^{23} centers per cubic meter. Referred to BaO, the activator concentration is thus a few parts per million. Concentrations of impurities as high as 0.001 per cent by weight will not seriously affect the cathode performance. As a rule of thumb, 1 microgram of poisoning material for every milligram of cathode coating may be considered to be enough to destroy cathode emission. These figures emphasize the necessity for great cleanliness, not only in the cathode materials, but in every part that enters the finished tube.

Every tube type must be considered as an individual within a species. Schedules for baking, evacuation, activation, and aging cannot usually be taken over bodily from one tube type to the next. Proper procedures for any new type must, therefore, be developed experimentally, and frequently with considerable difficulty.

CONCLUSIONS

In the field of oxide cathodes, perhaps more than in any other field, it is difficult to make many specific recommendations that will apply to all situations. Accordingly, the author has attempted to generalize and to present a broad picture of theory and practice. It is hoped that the presentation is broad enough to permit the reader to particularize according to his own needs.

LIST OF SYMBOLS

n_c = density of electrons in the conduction band.
N_c = density of available states in the conduction band.
q = electron charge.
ζ = energy of the Fermi-level measured from the bottom of the conduction band.
k = Boltzmann's constant.

T = absolute temperature.
h = Planck's constant.
m = mass of an electron.
N_i = density of states at the energy E_i
E_i = energy of a state in the forbidden band (measured from the bottom of the conduction band).
n_{oi} = the density of the electrons in the i^{th} state at a temperature of absolute zero.
σ = electrical conductivity of the oxide coating.
μ = mobility of the electron in the conduction band.
j_o = thermionic emission current density with zero-field at the cathode surface.
χ = electron affinity of the oxide-coated cathode.
N_b = density of impurity levels in the cathode coating.
E_b = energy of the impurity level (measured from the bottom of the conduction band).
ϕ_o = work-function of the cathode (as expressed in Richardson's equation).
V_a = anode voltage.
R_k = cathode coating resistance.
i_k = cathode current
ϕ_a = anode work-function
ϕ_c = cathode work-function
L = anode-cathode spacing.
$\vec{D}$ = electric displacement.
ρ = charge density.
v = velocity of the electron at the potential ψ.
ψ = potential within the cathode-anode space.
v_o = velocity of the electron at zero potential.
$\vec{j}$ = electron current density (vector).
n = density of the electrons in the cathode-anode space.
$\vec{v}$ = velocity of an electron in the cathode-anode space
ϵ_o = permittivity of free space.
$\vec{E}$ = electric field
j_1 = current density under idealized space-charge theory.
j_2 = current density under practical space-charge conditions
V_s = height of the space charge barrier.
l = distance of the top of the space charge barrier from the cathode surface.
V = vacuum gap potential.
j = electron current density (scalar).
E = the electric field at the surface of the cathode.
j_3 = current density under accelerating conditions.

REFERENCES

1. Wehnelt, A., S.-B. physik.-med. Soz., Erlangen, 95, 115, 1903
2. Wehnelt, A., Verh. Dtsch. physik., Ges., 5, 346, 1903
3. Wehnelt, A., Ann. Physik 14, 425, "On the Emergence of Negative Ions from Hot Metallic Compounds", 1904
4. Wilson, H. A., Proc. Roy. Soc., Lond., A134, 277-287, 1931
5. Wagener, S. and G. Hermann, The Oxide Cathode, Vols. I and II "The Oxide-Coated Cathode," Chapman and Hall, 1951
6. Rittner, E. S., "A Theoretical Study of the Chemistry of the Oxide Cathode," Philips Research Reports, 8, 184, 1953
7. Becker, J. A., "Phenomena in Oxide-Coated Filaments," Phys. Rev., 34, 1323, 1929
8. Nergaard, L. S., "Studies of the Oxide Cathode," RCA Review, 13, 464, 1952
9. Plumlee, R. H., "The Electron Donor Centers in the Oxide Cathode," RCA Review, 17, 231, 1956
10. Private Communication
11. Nottingham, W. B., "Thermionic Emission," Handbuch der physik, Vol. XXI, Springer, Berlin, 1959, Academic Press, New York, 1948
12. Eisenstein, A. S., "Oxide-Coated Cathodes," Advances in Electronics, Vol. 1, p. 1, 1948
13. Schottky, W., Physik Z., 15, 872, 1914
14. Benjamin, M., and H. P. Rooksby, "Emission from Oxide-Coated Cathodes," Phil. Mag., 15, 810, 1933
15. R. Loosjes and H. J. Vink, "The Conduction Mechanism in Oxide-Coated Cathodes," Philips Res. Rep., 4, 449, 1949

Electron-Emission Coating for the Oxide Cathode

C. H. Meltzer and E. G. Widell

Harrison

BASIC CONSIDERATIONS FOR THE EMISSION COATING

The oxide-coated cathode is a system for producing efficient emission of electrons by the application of thermal energy to a material having a low work function. The system uses a nickel alloy for the cathode substrate metal upon which the oxide coating derived from alkaline-earth carbonates is formed. The presence of an activated oxide coating lowers the work function of the substrate nickel core metal. The low work function of the system permits operation at the relatively low temperature of 1025 K, and copious electron emission at low heater-wattage input. Electron emission is influenced by the effects of the low work function as well as by the electron transfer efficiency involved in the porous nature of the emission-oxide matrix coating.

Work Function

The work function ϕ, or the total amount of work necessary to free an electron from a solid, is measured in electron-volts (ev) and is temperature dependent. The temperature dependence may be produced either by the work function of the base metal, or by absorbed layers of atomic films. In general, the work function increases with the decrease in atomic volume or lattice spacing, i.e., the work function increases with the increasing electron affinity by reason of increasing electrical forces of attraction to the nucleus with decreasing distance, as shown in Fig. 1. When atoms of a lower work function are adsorbed onto the surface of a higher work function metal, the work function of the substrate metal is reduced by the potential barrier at the surface of the metal. For example, when a monolayer of barium (ϕ = 2.52 ev) is vaporized onto the surface of nickel (ϕ = 4.96 ev), the resultant adsorption layer, barium-oxygen-nickel, has a work function of 0.9 ev.[1]

Although large lattice spacings in a solid contribute to a low work function, they are also associated with low melting points and low boiling points. Therefore, the usefulness of materials having large lattice spacings is limited by their high rate of evaporation, which tends to shorten the life of a system under vacuum tube conditions as well as to contaminate adjoining electrode surfaces. A balance of the two factors — work function and rate of evaporation — is desired in an oxide-cathode system. The barium/barium oxide system has the lowest usable work function consistent with minimum rate of evaporation and long life at the operating temperature of the tube. Fig. 2 illustrates the relative emission capability of the oxide, tungsten, and thoriated-tungsten cathode systems.

The work function of a metal not only involves the particular substrate metal but also involves the type of crystal surface that is being measured. In some cases, preferred orientation occurs and it is found that the work functions of different faces of the metal lattice will vary with the plane of the lattice that is exposed.[2,3] However, for practical purposes, the values of the work function assigned to those materials used for cathode systems apply to polycrystalline surfaces with random orientation such that average effects are created.

Semiconductor Plus "Electron-Pore Gas" Conductivity

The activated alkaline-earth oxide coating on a cathode nickel base metal lowers the work function of the system by modifying the surface-energy potential barrier of the substrate metal so that electrons are transferred copiously at a relatively low operating temperature. In addition to creating a low work function system, which can be obtained by a thin film of oxide coating, a relatively thick oxide coating (0.75 to 3.5 mils) introduces a factor of coating porosity that allows an "electron pore gas"[4,5,6] effect to operate in conjunction with the basic semiconductor mechanism of electron transfer.[7,8,9] Normally, pure oxide coating has a resistance of 10^{12} ohms. The activating process causes the oxide to act as a semiconductor by a process of physical-chemical reactions which are believed to lead to the formation of barium and associated donor sites in the oxide lattice. In addition to this semiconductor property of electron transfer, there exists a phenomenon of electron transfer through an "electron pore gas" in the interstices between the macroscopic crystal particles of the oxide matrix at temperatures above 575 K. Although the electron gas, i.e., a cloud of electrons which is considered in terms of kinetic theory of gases, does not reduce the work function of the system, it does increase the efficiency of emission by increasing the conductivity of the system, as shown in Fig. 3.[10] A higher level of emission performance is possible at the operating temperature of 1025 K because there is a greater net voltage applied between cathode and plate as a result of the lower voltage drop across the oxide coating, i.e., the depth of the poten-

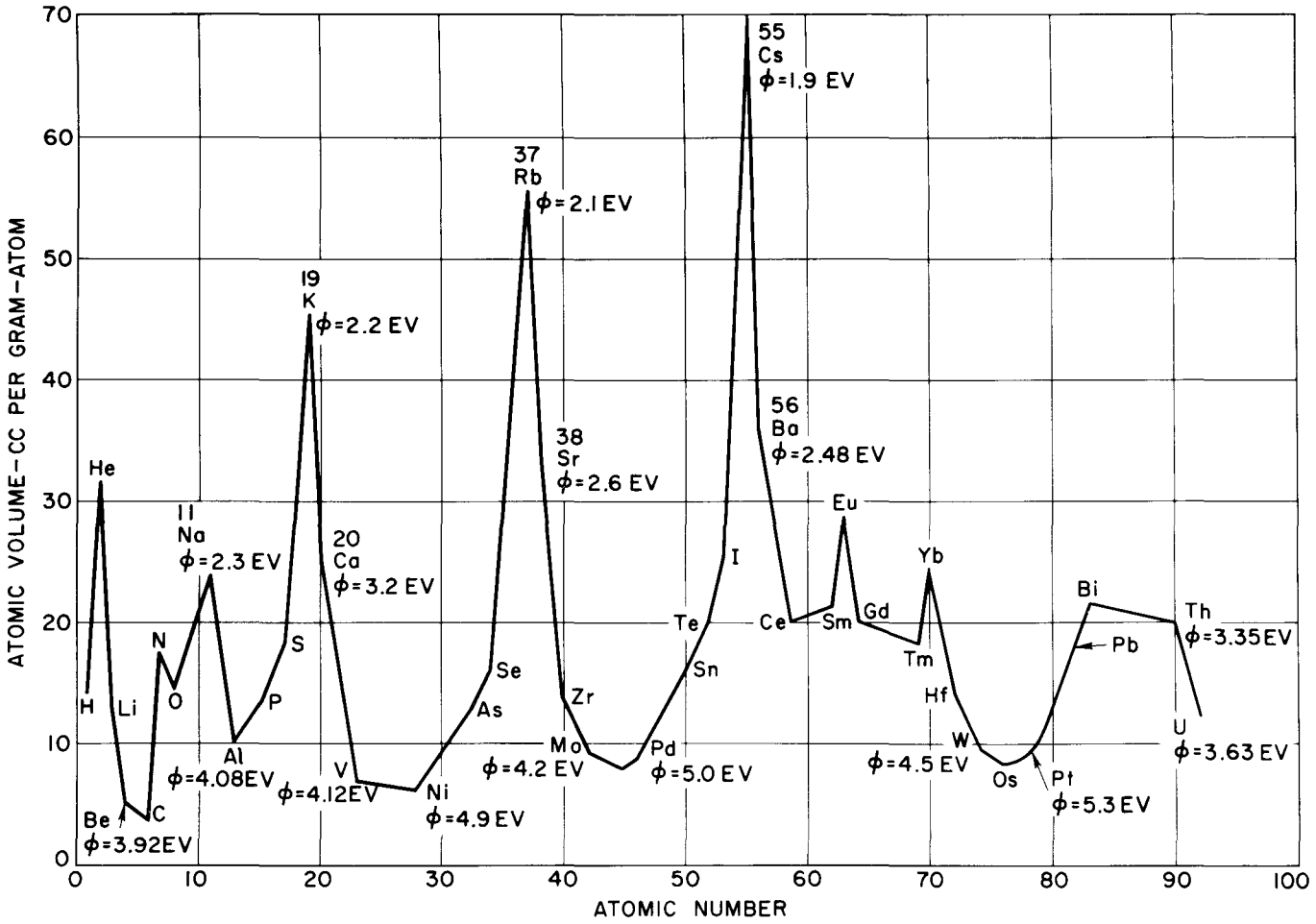


Figure 1. Atomic Number Versus Atomic Volume for Elements with Various Work Functions (Ø)

tial minimum, as well as its distance from the surface of the cathode, is decreased to permit more of the electrons to reach the plate. A relatively thick porous oxide coating, in addition to contributing a higher conductivity by reason of the electron pore gas effect, contributes to long life performance because the matrix of barium oxide in the coating is a reservoir for active barium sites in the system by reason of the chemical reduction processes induced by the active reducing agents of the nickel base metal.

Activation of the Oxide-Cathode System

Barium and associated donor centers are produced by chemical reduction of the barium oxide by reducing agents in the cathode nickel-alloy base metal. Barium oxide plus metal yields the metal oxide plus barium. Barium oxide can not be decomposed to barium and oxygen within the temperature limitations of the vacuum tube. The temperature required to decompose barium oxide would be close to 3000 K at a vacuum of 10^{-5} millimeters of mercury. Electrolytic dissociation of barium oxide yields only about two per cent of the available barium centers. Electrolysis of barium oxide occurs upon passage of current through the coating whereupon the barium ion is transported toward the base metal to be reduced to free barium by acquiring two electrons. The oxygen ion forms atomic oxygen yielding two electrons in the process.[11] Oxygen ultimately leaves the oxide coating to enter the vacuum space where, in time, it is removed from the vapor phase by combining with the getter flash material. As soon as barium has been created and donor sites have been established in the oxide matrix, the oxide coating is capable of transferring high emission current. Electrolysis does not activate an oxide system as readily as chemical reduction by certain alloying metals in commercial cathode nickel. The time for activation by electrolysis is measured in tens of hours for a pure platinum cathode in contrast to a time of minutes for activation by chemical reduction, as shown in Fig. 4.

Electrical Importance of the Substrate Metal and Emission Coating

The reducing-element content of the cathode nickel base metal creates and replenishes the supply of barium centers which maintain the low work function of the system and the basic semiconductor method of electron transfer. The operating characteristics of the cathode with respect to other tube electrodes are determined by composition, purity, and particle size of

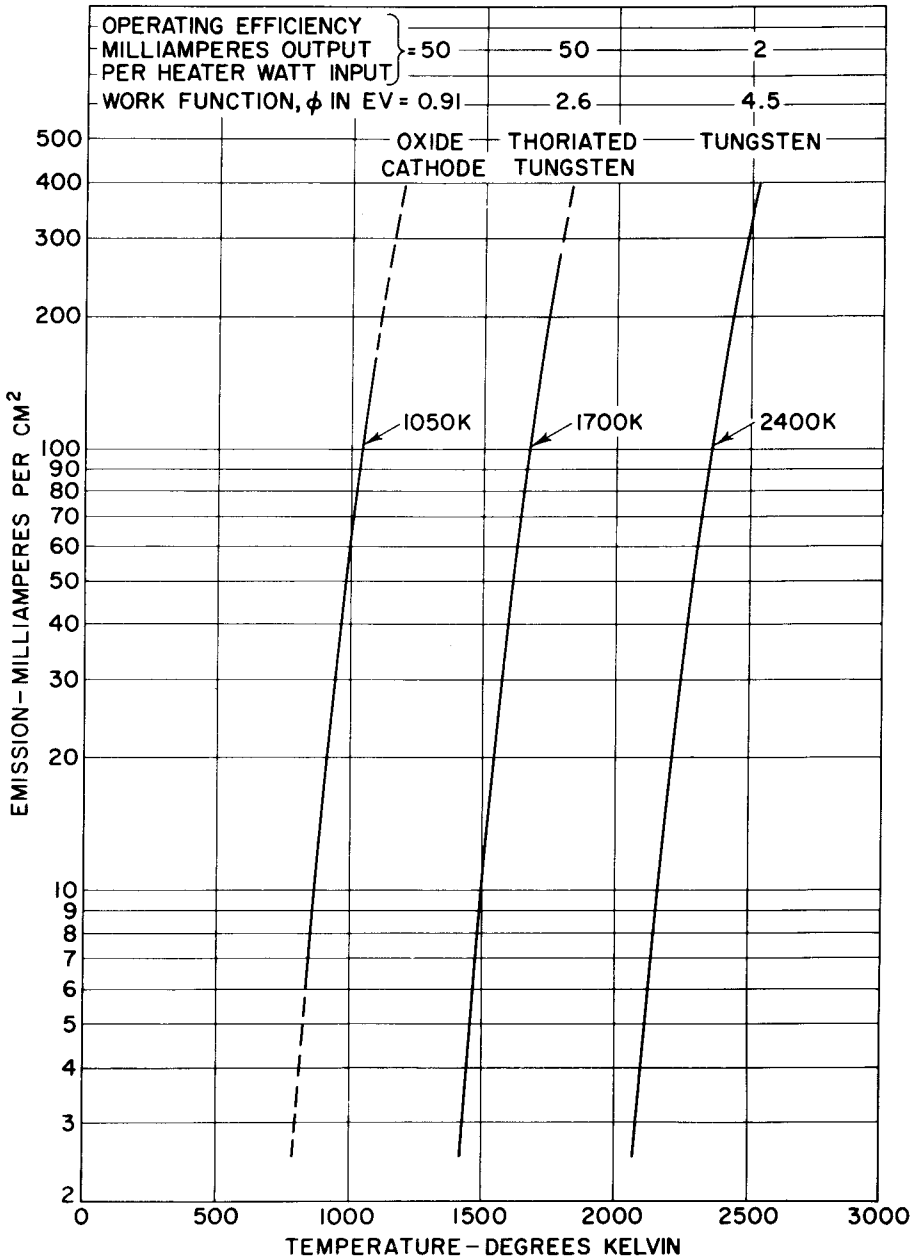


Figure 2. Average Value of DC Emission Current from Different Cathode Systems

the emission carbonates; the technique of coating the cathode metal; the conversion of the carbonates to the oxide-matrix crystal structure; the rate of thermal diffusion of barium through the coating; the rate of electrolytic transport of the barium ion; the rate of evaporation of barium/barium oxide;[12, 13, 14] and the concurrent formation of surface films on adjoining elements. The location of the active barium donor sites—whether in, on, or around the crystal lattice of the alkaline-earth oxide—is not known with certainty. The electrical importance of the crystal structure of the oxide lattice is related to the degree of porosity, which determines the effectiveness of the electron pore gas in decreasing the resistance of the coating at the operating temperature of 1025 K as shown in Fig. 5.

Equilibrium Systems and Interactions

In addition to the variable factors noted above, further variations on the physical-chemical equilibria which influence the micro-chemical actions occurring in the oxide coating are induced by the effects of temperature, time for decomposition of the carbonate to the oxide, coating weight, particle size, and density. Evaporation of certain condensable metals from the cathode base metal, as well as from the oxide coating, tends to contaminate insulating mica spacers as well as the heater insulation coatings. The effect of temperature, time, and partial pressure of carbon dioxide and carbon monoxide gases influences the porosity of the coating in terms of sintering and fusion of the coating. Elec-

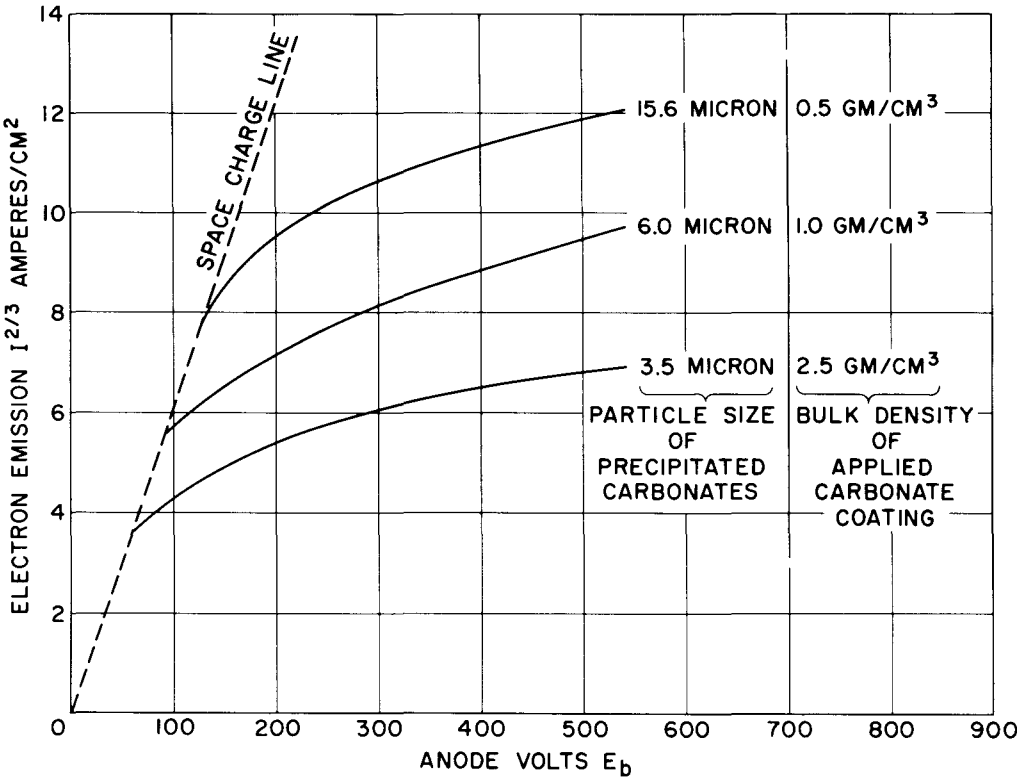


Figure 3. Pulse-Emission Performance from Oxide-Coated Cathodes Using the Same Cathode Base-Metal Nickel Alloy, but with Oxide Coatings Derived from Carbonates of Different Particle Size as Originally Precipitated. Particle-size determination by the Andreasen pipette sedimentation method. Type 1-v tube with E_f = 5.5 volts; plate voltage of one-microsecond duration at a pulsed repetition frequency of 1000 cycles per second.

tron bombardment of electrode surfaces and glass walls of the tube releases gaseous products that interact with the active barium centers of the coating causing a reduction of emission; positive-ion bombardment can disrupt surface layers of the oxide cathode to cause an increase in the work function of the system. In view of the varied factors and phenomena involved,[15,16] it may be concluded that all equilibria in the oxide cathode are in delicate balance with each other in the sequence of activation and regeneration during the life of the tube. These equilibria are dependent upon the non-injurious nature of the environmental vapor phase for the best electron-transfer performance and long life.

PREPARATION OF THE EMISSION COATING

The Carbonate Form

The preparation of the alkaline-earth emission carbonates is a critical operation in terms of maintaining purity, uniformity, particle size, and chemical composition. The carbonates are formed by precipitation in which a solution of sodium and/or ammonium carbonate is added to a solution of alkaline-earth nitrates. The solutions are made under controlled conditions using distilled water and are filtered before being mixed together in the reaction. The reaction is of the double decomposition type in which the equilibrium is shifted in the forward direction because of the insolubility of the alkaline-earth carbonate which precipitates out of solution. For example,

$$\underset{\text{barium nitrate}}{Ba(NO_3)_2} + \underset{\text{sodium carbonate}}{Na_2CO_3} \rightleftharpoons \underset{\text{barium carbonate}}{BaCO_3}\downarrow + \underset{\text{sodium nitrate}}{2NaNO_3}$$

$$\underset{\text{barium nitrate}}{Ba(NO_3)_2} + \underset{\text{ammonium carbonate}}{(NH_4)_2CO_3} \rightleftharpoons \underset{\text{barium carbonate}}{BaCO_3}\downarrow + \underset{\text{ammonium nitrate}}{2NH_4NO_3}$$

Various precipitation methods can be used, but to obtain reproducible uniform particle size distribution all methods must adhere to standardized control procedures with respect to concentration of reacting solutions, temperatures, rate of addition, rate of mixing, and pH of the reacting media. The carbonates that are used for emission purposes are usually the coprecipitated double carbonate of barium and strontium in 50/50 mole per cent composition (57/43 weight per cent) or the coprecipitated triple carbonate of barium, strontium, and calcium in 57/39/4 weight per cent composition. The carbonates may be made as the needle form (orthorhombic crystal lattice) or the spherulite form (rhombohedral crystal lattice) depending upon the temperature, composition, and concentration of the reactant solutions. The carbonate particles as precipitated in needle or spherulite form are composed of crystallites representing a lattice arrangement. The carbonate crystal particles are from 1 to 25 microns long, and from 1 to 4 microns in diameter. Crystallite sizes are from

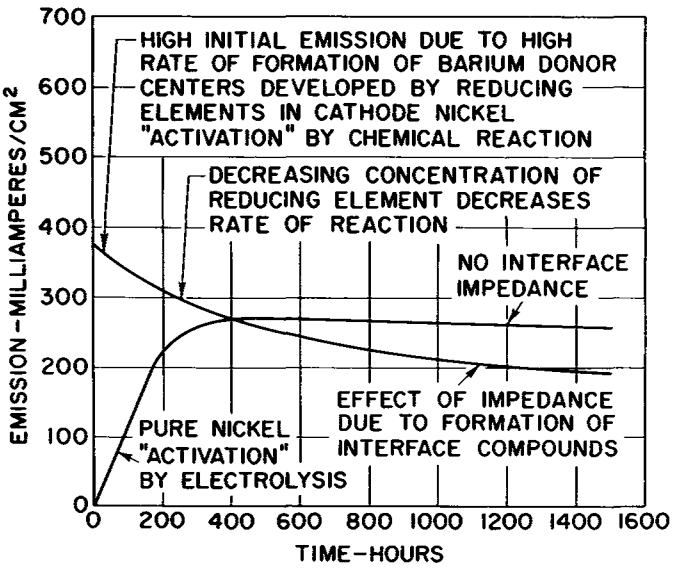


Figure 4. Difference in Level of Emission Obtained from Oxide-Coated Cathodes Using the Same Oxide Coating but Different Cathode Base-Metal Alloys: (1) a reducing-element-bearing nickel alloy having a high rate of initial activation by chemical-reduction processes, but also, having increased impedance effects due to interface-compound formation; (2) a pure nickel alloy having a slow rate of activation by means of electrolysis action, but then maintaining a relatively stable level of performance on life.

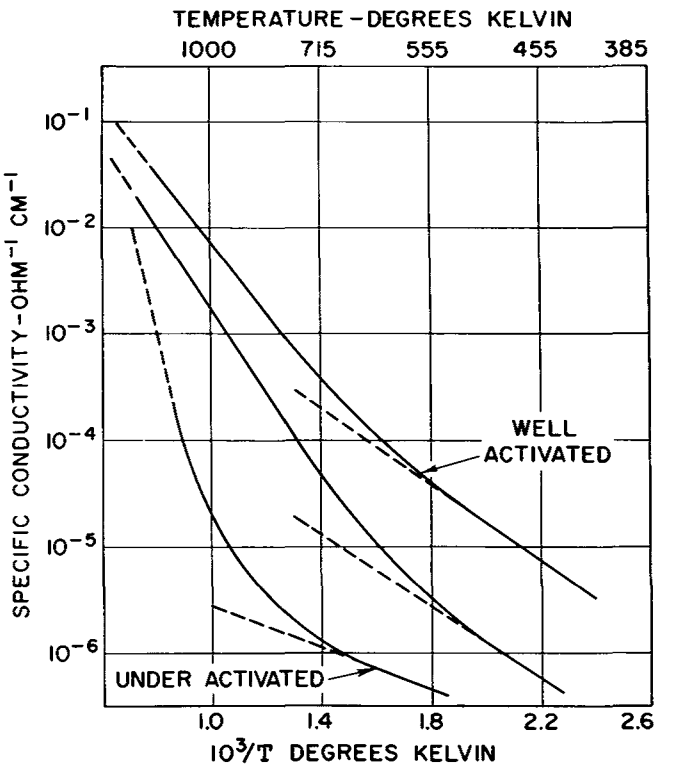


Figure 5. Conductivity of Oxide Coatings as a Function of Temperature for Different Degrees of Activation and Pore-Gas Conduction

100 to 1000 angstroms. Figs. 6 and 7 show representative needle shaped crystal of triple carbonates precipated by sodium and ammonium reagents.

The methods of precipitation are essentially straightforward. However, because purity of the product is essential to emission performance, the effects of adsorption and entrapment of foreign ions are minimized by the use of proper sequences of precipitation and washing operations. There are two precipitation methods: (1) a batch method wherein one reagent solution is added to the other reagent solution and (2) a continuous flow method wherein both reagent solutions are brought confluently into a reaction zone.[17] In the batch method, the addition of alkali carbonate solution to the alkaline-earth nitrate solution is preferred to the addition of the nitrate solution to the carbonate solution because the former procedure insures the production of a chemically pure product with a smaller amount of contamination from the alkali ions and salts. In the second method, the simultaneous addition of equimolar proportions of alkaline-earth nitrates and alkali carbonates minimizes the impurity content of the adsorbed alkali nitrate ions because the concentration of the alkali nitrate ions is at a minimum at the time of the formation of the alkaline-earth carbonate crystal. Another important feature of continuous flow methods of precipitation is the greater uniformity of the precipitated particle size than that obtained by the batch process. The continuous-flow reaction approaches one of constant-volume constant-con-

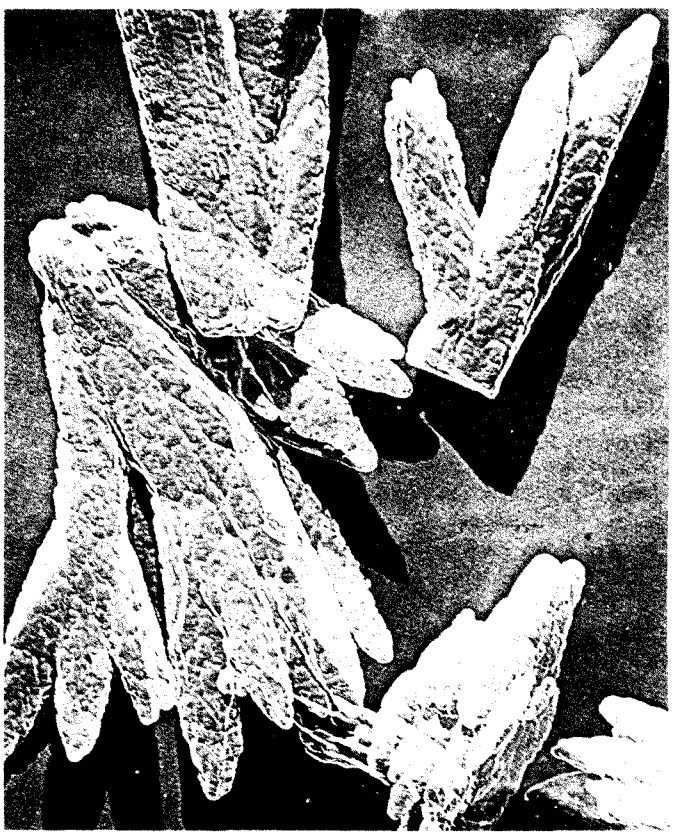


Figure 6. Sodium-Precipitated Triple Carbonate Photographed by Carbon-Replica Technique (electron-microscope magnification x 50,000; courtesy Dr. E.P. Bertin)

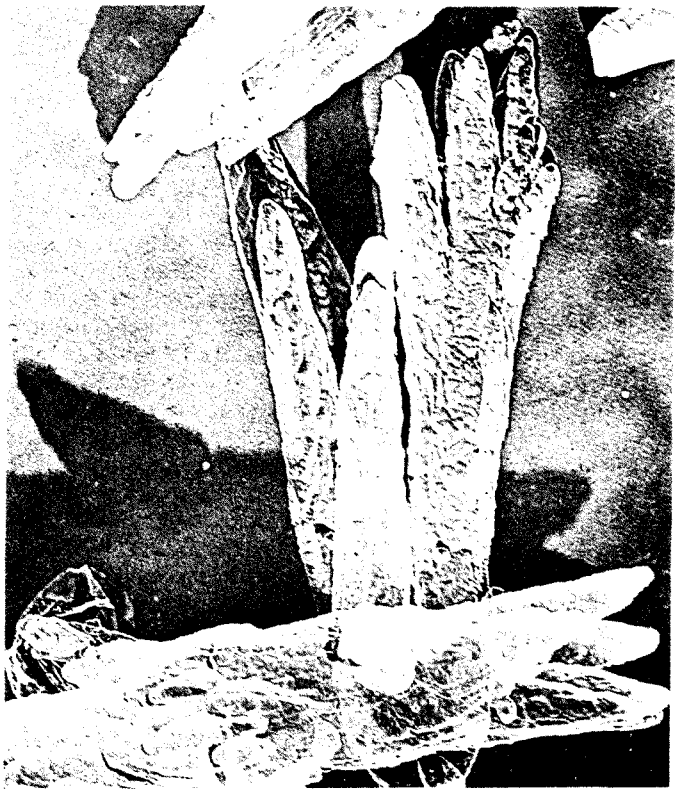


Figure 7. Ammonium-Precipitated Triple Carbonate Photographed by Carbon-Replica Technique (electron-microscope magnification x 50,000; courtesy Dr. E. P. Bertin)

centration during precipitation instead of undergoing dilution as does the batch type of reaction.

Emission carbonates precipitated by the addition of sodium carbonate differ from those prepared by ammonium carbonate reaction in particle size and residual impurity content of the alkali salt.[18] Under equivalent conditions, the "ammonium" type of emission carbonate will tend to be about 10 per cent larger than the corresponding "sodium" type of emission carbonate when precipitated in the temperature range above 80 C. The difference of impurity content between "ammonium" and "sodium" types of emission carbonates is that the former is ammonium nitrate and the latter is sodium nitrate. Both ammonium nitrate and sodium nitrate are adsorbed to an equal extent onto the surfaces of the carbonate crystals or entrapped within the crystals. Both nitrates are washed from these carbonates to the same extent. The difference between ammonium nitrate and sodium nitrate is in the difference in melting points and decomposition products. Ammonium nitrate melts at 169 C and decomposes into gaseous nitrogen, oxygen, and water vapor at 210 C; whereas, sodium nitrate melts at 308 C and decomposes into sodium oxide, nitrous oxide, and oxygen at 380 C. Thus, when the residual impurity is trace amounts of ammonium nitrate, it is lost during bakeout or sealing-in operations. During decomposition from the carbonate, the emission oxide coating is not contaminated with any low-melting-point substances that might cause excessive sintering and fusing of the coating. Where the impurity is sodium nitrate, the higher melting point and decomposition temperature contribute to side effects in the early decomposition stages of the carbonates leaving a residue of sodium oxides which sublimes at temperatures above 800 C. The low melting point of the sodium salts contributes to the inherent melt condition of the eutectic of barium carbonate ($2BaCO_3$:1BaO) which melts at 1030 C. The over-all effect is to increase the tendency toward fused oxide coatings under high speed sealex operation where both the heater wattage and induction heating by RF coils are high, and vacuum pumping efficiency is low.

The particle size of the precipitated emission carbonate and the particle-size distribution of the carbonate are of importance in determining the ultimate porosity of the oxide coating. It is this porosity that contributes to the electron pore gas effect in decreasing the electrical resistance of the coating.[19,20,21] Within the limits of the controlling factors of concentration, temperature, pH, and rate of input mixing of the confluent reagent solutions, the method of continuous flow precipitation can give a narrower distribution of particle size range about a given desired particle size than the batch method.

The factors of temperature and concentration, pH of the reacting media, and rate of mixing of the reagent solutions are major determinants of the shape and size of the precipitated carbonate crystals. The crystals are produced as relatively large needles in the temperature range about 80 C. A mixture of crystal forms, varying from spatulate clusters of needles to spherulite forms, are obtained in the mid-temperature range from 50 to 75 C. Spherulite formation is predominant at temperatures below 50 C. All crystal forms within the composition ranges that are used for emission purposes are of the orthorhombic type of lattice formation. Particle size varies directly with temperature and inversely with concentration of the reacting solutions, i.e., needle size increases with increasing temperature; needle size decreases with increase of concentration, provided the temperature is constant. When the pH of the reacting media is below 7, the precipitated carbonate crystals are larger than those precipitated from a reaction medium above pH 7. The rate of addition of the reagent solutions to each other influences the particle size; rapid addition causes small particle-size formation, whereas slow addition causes large particle-size formation. The rate of mixing, the shape of the impeller blade, and the turbulent currents produced in the reactor vessel also influence the crystal habit of nucleation and growth. Manufacturing practice has resulted in the standardization of processes designed to control all these factors to the high degree necessary for the production of uniform, chemically pure, ultrafine quality emission carbonates.

The relationship between the size and shape of the emission carbonate particle as precipitated and the emission performance of an oxide coating is complex.[22,23] In general, carbonate particle sizes within the limits of 2 to 18 microns in length and 2 microns in diameter yield an oxide coating capable of a higher level of emission than can be obtained from carbonates

which are originally precipitated larger than 25 microns long or smaller than 2 microns long. Emission carbonate crystals that are precipitated either too large or too small yield oxide coatings that are not capable of good emission performance. Ideally, the larger the carbonate-particle size, the higher would be the level of emission obtainable from the oxide coating from these carbonates by reason of the greater degree of porosity. The deviation from this effect in the instance of very large precipitated particles results from three factors.

First, the growth of large carbonate crystals is brought about by long contact with the mother liquor wherein small crystals are solubilized and redeposited onto larger crystals. Such large crystals contain many crevices and cavities enclosing mother liquor solution. These occluded impurities are entrapped by a bridging of the fissure at the surface of the crystal. As a result, it is difficult to remove the impurities by washing. Coarse crystalline material can be much more heavily contaminated than the same material in a finer state of subdivision. However, when the particle size is too small, the effects of adsorption onto the increased surface area also creates a high degree of impurity which is very difficult to remove. Thus, a balance point in terms of optimum particle size must be made between too large or too small a particle size.

Second, large crystals require a longer or harder ball-milling to grind the particle down to dimensions usable for coating techniques calling for smoothness and density needed for close-spaced tube design or high voltage gradient characteristics. The prolonged milling necessary to reach a fine state of subdivision builds up the abrasive content from the ball-mill pebbles (a potassium aluminum silicate clay complex) and, thus, tends to increase the electrical resistance of the coating.

Third, the more porous type of emission coating derived from carbonates of large particle size is capable of being decomposed to the oxide more readily and at somewhat lower temperatures. However, the very porosity of the crystalline matrix subjects it to more extensive sintering from unfavorable gas equilibrium conditions that lead to eutectic-melt formation.

In contrast to these effects of large particle size carbonates, when the carbonate powders are precipitated as very small particles, they are difficult to process and handle in centrifugal washing and filtering operations; the adsorption effects resulting from the increased surface area of matter in a finer state of subdivision increases the impurity content of absorbed alkali salts; and the denser type of emission coating derived from small particle size carbonates are more difficult to decompose on sealex processes and yield a lower level of emission performance due to the increased resistance of the coating that results from the much reduced porosity. Fig. 3 shows that the level of emission performance for the less porous type of coating is definitely lower because of the decreased conductivity of the coating. The denser type of coating, however, is more resistant to "poisoning actions" that tend to decrease the amount of active barium centers of the oxide matrix. The denser type of oxide coating offers more dimensional stability with respect to grid-to-cathode spacing and, because of the decrease of surface area within the bulk of the less porous oxide matrix, less barium/barium oxide evaporates to contaminate adjoining tube electrodes. In production practice, a balance of factors including particle size and porosity, decomposition characteristics, resistance to poisoning, and level of emission performance determines the final selection of the optimum type of carbonate material to be used.

After precipitation, the carbonates are collected in a centrifuge, washed, reslurried in fresh distilled water, centrifuged, and rewashed to free the crystals of residual mother liquor salts and impurities. Next, the carbonate powder is dried to a very low moisture content, blended, and analyzed for purity and composition. Acceptable bulk powder is then stored as approved lots. Carbonate powder is withdrawn from stock upon demand, redried to insure a moisture content of less than 0.10 per cent, and then ball-milled. The technique of application and the end use of the carbonate coating dictate the ball-milling time, selection of binder, and solvents. The number, weight, and size of the ball-milling pebbles, the quantity of carbonate powder and binder solution, the speed of milling in terms of revolutions per minute and peripheral speed, and the viscosity of the charge are maintained at standard conditions to insure reproducibility and uniformity of the emission carbonate coating.

The emission performance of the oxide coating is influenced by the texture and porosity of the emission carbonate coating. The physical properties of texture and porosity of the carbonate coating are transferred to the oxide form despite the conversion from the orthorhombic lattice of the carbonate to the cubic lattice of the oxide crystals. When the emission coating is deposited onto the base metal, the solvents evaporate and leave the carbonates mechanically bonded to the metal with the nitrocellulose or methyl methacrylate binder. The proportion of the weight of the coating to its thickness depends upon the type of cathode and the tube type. The weight and thickness of the coating (i.e., the packing density) must be uniform and reproducible. This uniformity is especially important for maintaining uniform transconductance values.

The type of solvents used in the carbonate-spray formulations influences the bulk density (porosity) of the emission coating as it is applied to the base metal. Solvents having a high vapor pressure (low boiling point) evaporate rapidly and produce a porous, fluffy coating. Solvents having a low vapor pressure (high boiling point) evaporate slowly to leave a smooth, dense coating.

The spray-gun aperture settings control the amount of material deposited per application, whereas, spray-gun air pressure and distance from the cathode control the degree of porosity of the applied coating. High air pressures tend to deposit dry, fluffy coatings; low air pressures tend to deposit wetter, more dense coatings. A dry, porous coating is deposited onto the cathode when the cathode is further away from the gun nozzle; a denser deposit of coating material is obtained when the cathode base metal is closer to the gun nozzle.

Low humidity and high temperature increase the rate of evaporation of the solvents and produce a more porous coating. On the other hand, high humidity causes the nitrocellulose binder to set in a brittle fashion and thus contributes to imperfect adherence of the coating to the base metal. The emission coating is applied in an air conditioned environment at 50 per cent relative humidity and at a temperature of 72 F to insure uniform spray application. The dew point of the water vapor in the compressed air is maintained at minus 40 F to minimize the effects of moisture on the setting of the nitrocellulose binder. The rate of air flow through the spray booth is kept constant to maintain a constant rate of solvent evaporation in order to obtain uniform deposition of the coating. The ratio of solids to solvents and the viscosity of the coating formulation are maintained within close limits. Low specific gravity formulations tend to deposit a wet, dense coating; high specific gravity formulations tend to deposit a dry, porous coating.

Whereas the true density of the triple-carbonate emission powder used in production is about 4.0 grams per cubic centimeter, the range of apparent density of the coatings of emission carbonates as normally deposited onto the cathodes in 0.50 to 2.0. Thus, the porosity of the carbonate coatings ranges from 87.5 to 50 per cent. The range of weight distribution of the carbonate coatings is 3.5 to 8.0 milligrams per square centimeter. A coating that is packed too densely has a lower level of emission performance than a fluffy, porous coating; however, when the coating is too fluffy, its mechanical adherence to the base metal is poor and the coating tends to powder. The standardizing notices show representative curves for weight, outside diameter, and apparent densities that can be used as a guide for design purposes. Ref. 24 describes coatings for oxide cathodes. The character of the spray is classified into three grades — wet, medium, and dry. In general, the adherence of the sprayed coating increases with the wetness of the spray during application. A wet spray produces a smooth, hard coating that is strongly adherent to the base metal and is difficult to damage in handling. A medium spray produces a coating that is visibly rough and which adheres firmly enough to withstand normal handling. A dry spray produces a coating that is decidedly rough and which must be handled carefully because of its fragility.

The use of the tube should determine the density of the coating. Close-spaced rectifiers, where arcing is a factor, require a high-density, smooth coating, as do cathode ray tubes and tubes with very close spacing. High emission performance can be better obtained with low-density porous coatings but only at the cost of lower resistance to poisoning and greater instability in performance. High-density coatings are more stable but have a lower level of emission performance; the quantity-rate of evaporation of barium/barium oxide is reduced to a minimum and grid contamination is less critical. In practice, compromises in coating weights and densities are required to obtain the desired tube characteristics under sealex conditions consistent with the base metal reducing activity and the temperature-time-pressure equilibrium during the decomposition of the carbonate coating.

The Oxide Form

Oxide-coated cathodes are used under a wide variety of conditions, and some of the requirements which come first to mind include such objectives as: low cost of preparation, high heater efficiency, long life at a continuous emission level of 80 to 130 milliamperes per square centimeter (Fig. 8), and reliable performance under pulse conditions of 40 to 80 amperes[25,26,27] per square centimeter (Fig. 9).[28] Tubes can be designed and built with oxide-coated cathodes to last more than 100,000 hours (about 10 years) by operating the cathodes at reduced temperatures. However, the lower the cathode temperature, the more likely is the oxide coating to yield lower emission due to the effects of effects of gaseous residues in the tube. This tendency of the oxide-coated cathode to "slump" under low temperature conditions results from the decrease in the rate of chemical reaction between the reducing element content of the base metal and the oxide. Sufficient active barium centers are not produced to offset the inactivation effects of those physical-chemical reactions from the environmental gas and vapor phase that tend to decrease the amount of active barium donors in the oxide matrix. When the oxide cathode is examined in terms of economy, ease of preparation and fabrication, relative mechanical stability, ease of activation, and relative low rate of evaporation consistent with long life, it is apparent that the alkaline-earth oxides have the lowest usable work function consistent with the objectives stated at the beginning of this paragraph.

Numerous methods are used for the application of the alkaline-earth compounds to the cathode base-metal.

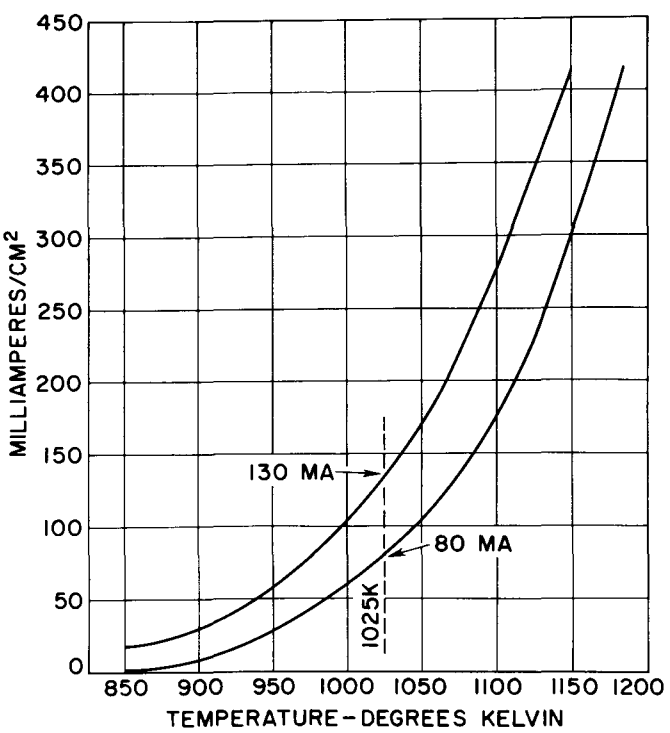


Figure 8. Permissible Direct-Current Density (ma/cm²) for Long Life Performance for Average Oxide-Coated Cathodes

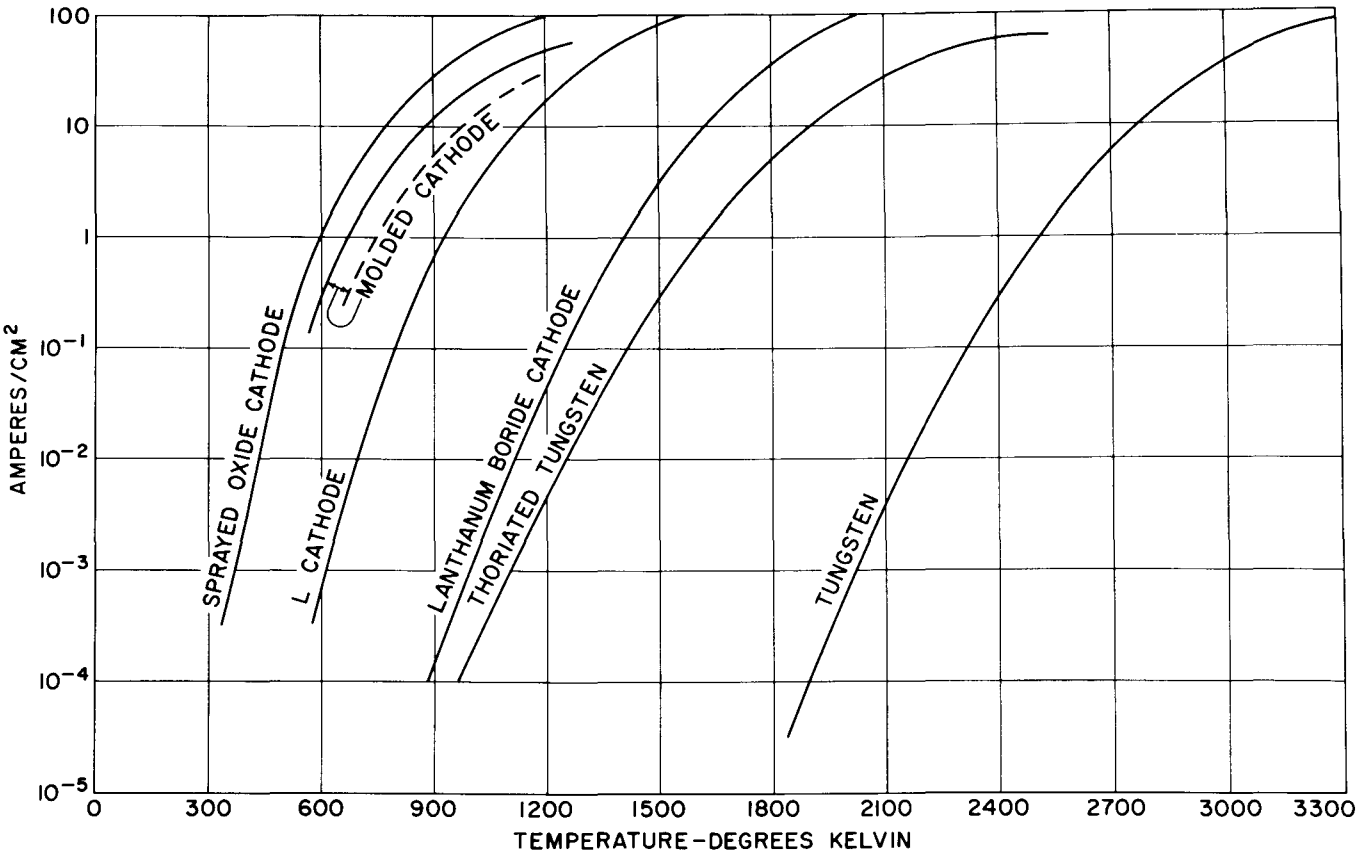


Figure 9. Pulsed-Emission Levels of Cathode Systems[28]

The emission coating can not be applied directly in the form of the oxide during the manufacturing operation because it reacts with water vapor and carbon dioxide in the air to form the hydroxide and carbonate compounds.

$$BaO + H_2O \rightarrow Ba(OH)_2$$

$$BaO + CO_2 \rightarrow BaCO_3$$

Therefore, some barium compound that is stable under atmospheric conditions, pure, and easily convertible to the oxide under vacuum conditions is required for coating the cathode base metal. The alkaline-earth carbonates have this required stability, purity, and convertibility. Although the system barium/barium oxide has the lowest usable work function of the family group (barium, strontium, calcium) at the operating temperature (the work function, in electron volts, for Ba/BaO on nickel is 1.1 at 1025 K; for Sr/SrO on nickel, it is 1.4 at 1175 K; for Ca/CaO, it is 1.9 at 1300 K), the barium/barium oxide system is not useful as an emission coating when used alone, because, as shown in Fig. 10, it inherently passes through a comparatively low-melting eutectic phase while being decomposed from the barium carbonate to the barium oxide.[29,30] However, the strontium and calcium systems do not pass through such a eutectic phase during the conversion from the carbonate to the oxide form at these temperatures:

$$3BaCO_3 \rightleftharpoons \underset{\text{(eutectic)}}{2BaCO_3{:}1BaO} + CO_2\uparrow \rightleftharpoons 3BaO + 2CO_2\uparrow$$

$$3SrCO_3 \rightleftharpoons 3SrO + 3CO_2\uparrow$$

$$3CaCO_3 \rightleftharpoons 3CaO + 3CO_2\uparrow$$

The melt condition of the eutectic phase of barium carbonate/barium oxide, $2BaCO_3 : 1BaO$, (melting point 1300 K) compared to that of barium oxide, BaO, (melting point 2170 K) causes the emission coating that is formed from barium carbonate alone to sinter to a great degree under certain conditions of temperature and carbon dioxide gas equilibrium pressure encountered during sealex operation. In practice, decreased porosity (i.e., fusion) and varying degrees of poor coating adherence to the base metal, in terms of brittleness, result from a greater cohesion for itself as a melt than adhesion to the base metal where the coating is of some thickness other than that of a thin film. The level of emission is reduced in a sintered oxide coating because of the decrease of porosity and increase in the size of the crystallite lattice. The conductivity of such a fused coating is decreased as the effect of the electron pore gas type of electron transfer is diminished. Consequently, the coprecipitated double carbonate of barium-strontium (50/50 mole per cent) composition was used because it minimized the eutectic melt formation of the barium carbonate component as a result of the diluent effect exerted by the strontium carbonate at

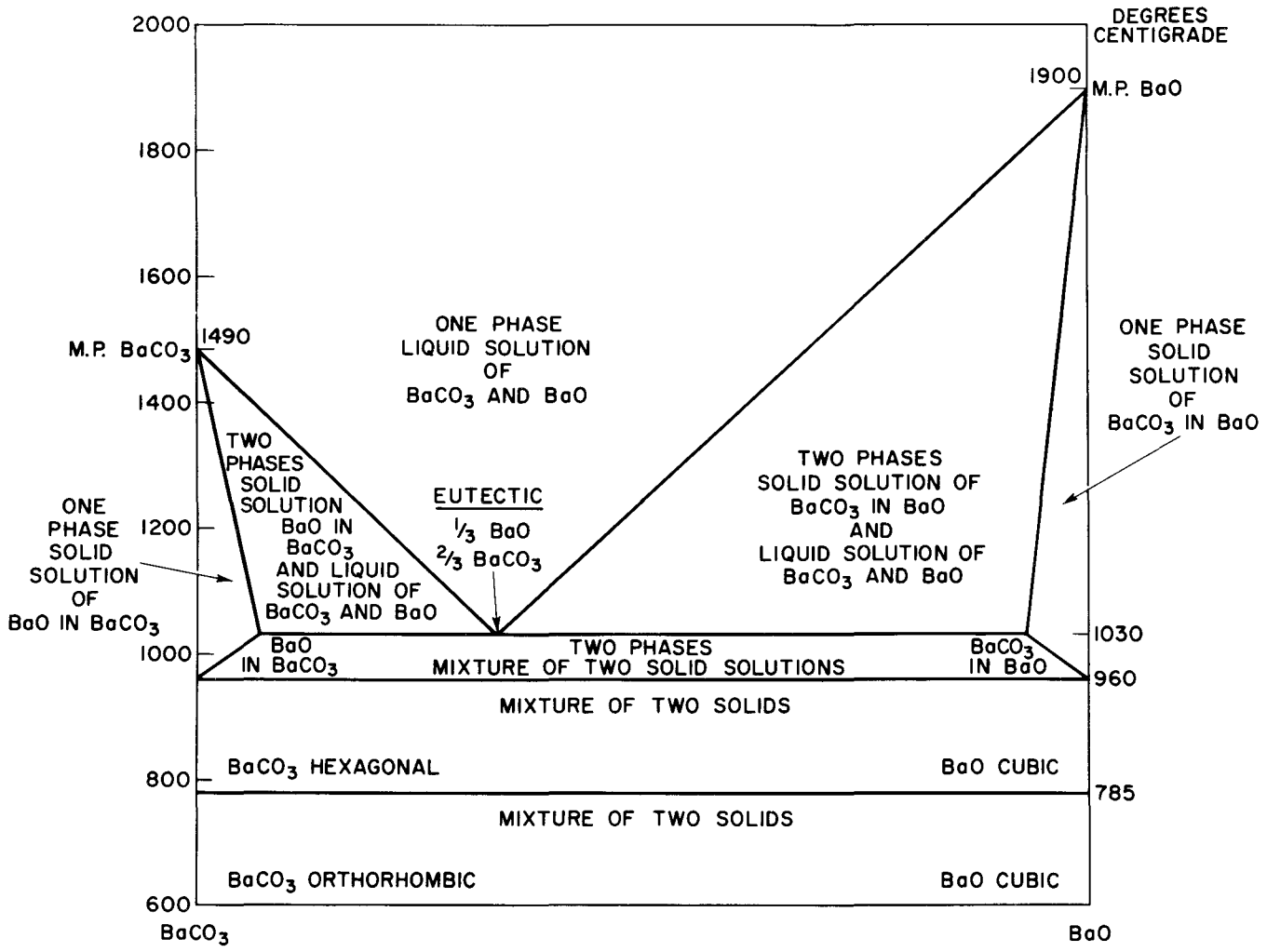


Figure 10. Phase Diagram for Barium Carbonate—Barium Oxide

the temperatures of processing and operation. In addition, the coprecipitated double carbonate, $(BaSr)CO_3$, yields larger crystals than the single barium or strontium carbonate under the same conditions of precipitation, i.e., standard temperature, concentration, and rate of reaction. The increased size of the precipitated particles, as carbonates, decreases the ultimate packing density of the emission oxide coating. Similarly, the use of the coprecipitated triple carbonates of barium, strontium, and calcium, $(BaSrCa)CO_3$ of 57/39/4 weight per cent composition, yields a still larger particle-size formation under the same conditions of precipitation. Figs. 6 and 7 show the large-size needle form of the crystal structure obtained for the triple carbonate. The conductivity of the oxide coatings derived from such carbonates is maintained at a higher level by virtue of the increased mean free path of the electron pore gas associated with increased porosity. The level of emission performance is consequently higher because of the lower voltage drop across the coating.

The complex changes that occur when the carbonate is converted to the oxide have been under investigation by many laboratories. The use of X-ray powder patterns as well as Geiger-counter display on graph recorders has shown that the crystallographic transformations that occur during the decomposition and subsequent heat treatment of the oxide are critical in determining the emission activity of the oxide coating. X-ray data have shown that for the coprecipitated crystal of barium-strontium carbonate $(BaSr)CO_3$,[31] there exists a range of lattice constants which vary continuously and linearly with composition indicating that true single-phase solid solution exists through the entire composition range. Similar studies on the triple carbonates $(BaSrCa)CO_3$[32, 33, 34] indicate that true solid solution can exist for the three component system; it is only when the calcium component is less than 10 mole per cent or 6 weight per cent of the composition ratio, as shown in Fig. 11, that true single-phase solid-solution exists.*

X-ray investigation of the internal crystallite formation of carbonate particles, by Eisenstein[35] has

* Bertin, E. P., C. H. Meltzer, and E. G. Widell, X-ray Study of the Triple Carbonates (unpublished)

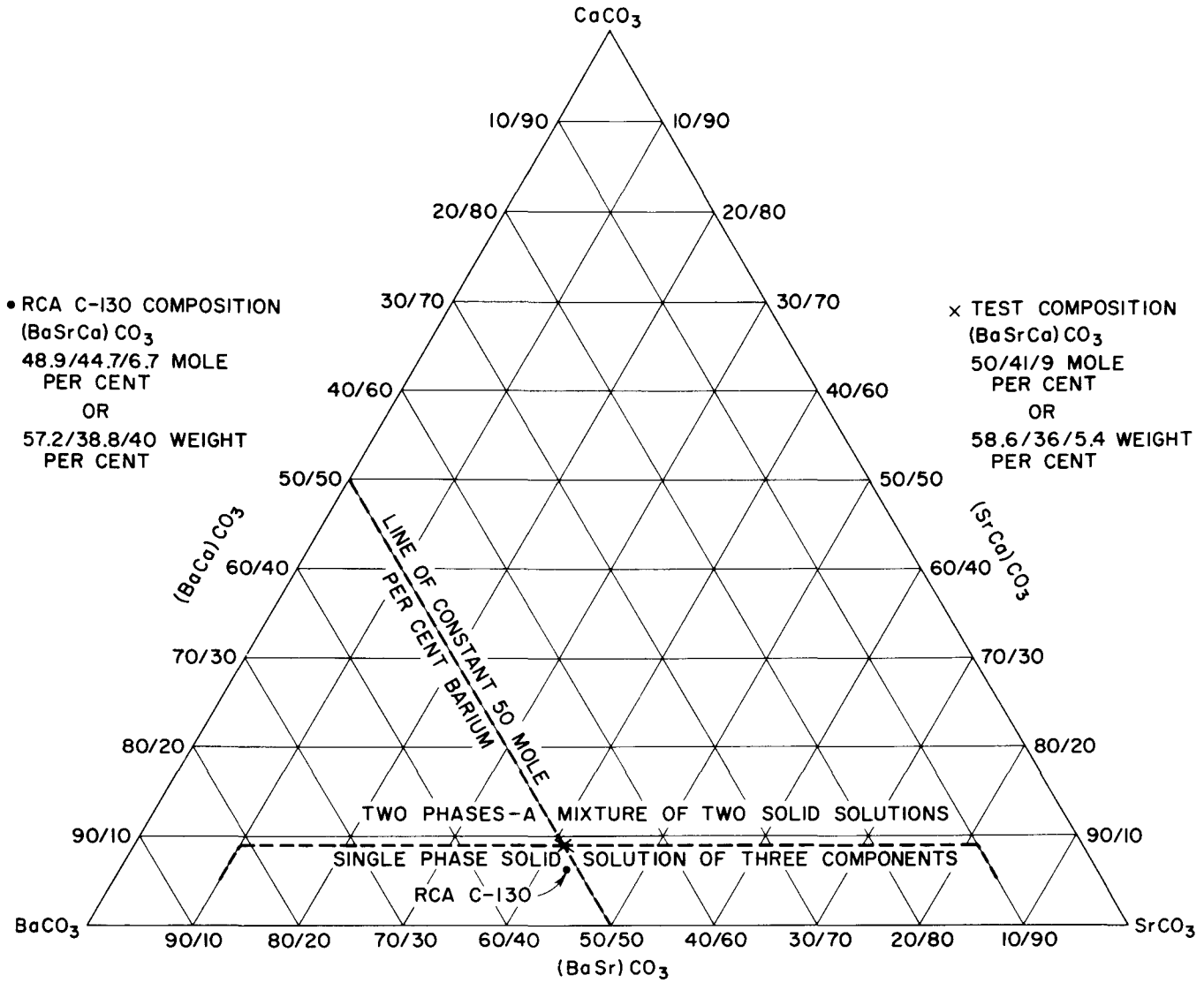


Figure 11. Ternary Diagram for Three-Component Solid-Solution Phase Systems of Barium—Strontium—Calcium Carbonates

contributed an important observation of the decomposition of the emission carbonates. It was noted that in the decomposition of the single-phase solid-solution double carbonate $(BaSr)CO_3$ crystals, the strontium carbonate component is converted to the oxide first; the barium carbonate component remaining in the lattice is converted some time later. When the decomposition is completed, the two oxides, SrO plus BaO, exist as separate phases; crystals of single-phase solid-solution (BaSr)O are obtained only after further heat treatment;[36] the critical temperature for this final transformation is about 1050 C. Similar decomposition phenomena occur for the triple carbonates of coprecipitated barium-strontium-calcium $(BaSrCa)CO_3$ within a limited range of composition when the calcium component is less than 6 weight per cent.

$$(BaSrCa)CO_3 \rightarrow \begin{cases} BaCO_3 \rightarrow BaO \\ SrCO_3 \rightarrow SrO \\ CaCO_3 \rightarrow CaO \end{cases} \rightarrow (BaSrCa)O$$

(orthorhombic) (hexagonal) (cubic) (cubic)

The important insight into this complex transformation is the fact that single-phase solid-solution carbonates ultimately yield oxides having small crystallite units of 150 angstroms, whereas, mixtures of the single carbonates, $BaCO_3$, $SrCO_3$, and $CaCO_3$, yield oxides having crystallite units of 250 to 300 angstroms, because of the higher temperature needed on heat treatment to create the single-phase solid-solution oxide from the separate, mechanically mixed, single carbonates.[35, 37, 38, 39] The single-phase solid-solution double carbonate $(BaSr)CO_3$ is of larger particle size than either of the separate single carbonates; the single-phase triple component solid-solution $(BaSrCa)CO_3$ (where calcium is less than 6 weight per cent) is larger than the comparable double carbonate when made under the same conditions. The coincidence of the resultant smaller crystallite formation in the oxide form with the porosity of the oxide coating derived from the larger-particle-size double or triple carbonate is related by a minimum fusion effect in explaining the higher emission. In a later paper, Eisenstein[40] concluded that the barium oxide molecules move to the surface of the coating by

diffusion to form crystallites and that the crystallites were the source of the low-work-function points from which emission can occur.

The electron-emission performance of the oxide coating derived from the carbonates of barium, strontium, and calcium is primarily correlated with the low work function of the system and its increased electrical conductivity as contributed in part by the porosity of the oxide matrix and the electron pore gas. The system barium/barium oxide has the lowest usable work function for emission purposes in vacuum tubes consistent with ease of activation and relative low rate of evaporation for long life. The presence of the strontium and calcium components in the triple carbonate composition contributes:

1. a diluent effect to offset the low-melting eutectic phase of the barium carbonate-to-barium oxide conversion.
2. a crystal habit of large particle growth of the carbonate form that determines the extent of the porosity of the oxide-matrix emission coating that is obtained.

The single-phase two-component solid-solution-oxide lattice exists for the double oxides (BaSr)O and (SrCa)O but not for barium and calcium because of the disparity of the size of their ions, i.e., barium possesses an ion radius of 1.35 angstroms, strontium has an ion radius of 1.13 angstroms, and calcium has an ion radius of 0.99 angstrom. The difference between the ion radii of barium and strontium is 16.3 per cent; whereas, the difference between the ion radii of barium and calcium is 26.7 per cent. Solid solution formation in the oxide form can exist only when differences between ion radii are within 17 per cent. X-ray data indicate that all the alkaline-earth-carbonate crystals that are made by coprecipitation to form $(BaSr)CO_3$, $(SrCa)CO_3$ and $(BaCa)CO_3$ consist of solid solutions throughout the range of composition (although not always single-phase solid solutions). However, only the oxides (BaSr)O and (SrCa)O become solid solutions upon conversion from the carbonates. The decomposition of $(BaCa)CO_3$ results in a matrix of a microcrystalline mechanical mixture of BaO plus CaO forms. In terms of significance of the effect of solid solution upon a favorable level of emission, the absence of the solid-solution formation for the BaO plus CaO matrix is important in the considerations to be given to too high a calcium content in the triple-component-oxide system that is derived from a triple carbonate orthorhombic (needle form) lattice structure.

Under standard conditions of preparation in which concentration, temperature, pH, rate of addition and rate of mixing are held constant, a series of double carbonates of barium and strontium compositions show an increase in the particle size of the precipitated carbonate crystal from 100 per cent barium to 75/25 mole per cent $(BaSr)CO_3$; the particle-size range and distribution remain essentially constant through the composition range from 75/25 to 25/75 mole per cent $(BaSr)CO_3$; as the composition ratio becomes richer in the strontium component the particle size decreases from the composition ratios 25/75 mole per cent $(BaSr)CO_3$ to 100 per cent strontium carbonate. Commercial practice undertook the preparation and use of the 50/50 mole per cent barium-strontium composition and controlled the particle-size range and distribution in order to influence the porosity of the final oxide-coating form that was used on the cathode base metal. The use of the 50 mole per cent barium component was undertaken for reasons of useful long life performance in terms of sufficient active barium in the emission coating for self-gettering, minimum contamination of adjoining electrodes due to evaporation of barium/barium oxide, and the above-mentioned porosity factor that controlled the conductivity of the oxide coating and, thus, influenced the level of emission performance of the tube.

The use of the coprecipitated triple carbonate of barium, strontium, and calcium was in commercial production in the early 1930 period. At RCA, the composition change was made by the addition of the calcium component at the expense of the strontium component while keeping the barium component constant in the system. Thus, the use of the former double carbonate of 50/50 mole per cent, or 57.2/42.8 weight per cent, $(BaSr)CO_3$ was discontinued in favor of the triple carbonate of 57.2/38.8/4.0 weight per cent $(BaSrCa)CO_3$ because of the increased level of emission performance that was obtainable from the use of such triple carbonnates as they were of larger particle size than the double carbonates when made under the same standard conditions of manufacture.

In discussing particle-size range and distribution of the emission carbonates, comparisons are made within a given series of compositions that are produced under standard conditions. Thus, in terms of tube performance as influenced by the porosity of the oxide coating, i.e., its conductivity, it is recognized that the greater porosity made possible by a change in the precipitation process (use of more dilute reactant solutions, or use of a slower rate of addition) is equally effective as the same degree of porosity achieved at a different concentration of reactants and with a change of component composition provided that the amount of the barium component remains at a value that is useful for tube performance. Precipitated carbonates can be made with the same particle-size range and distribution at two entirely different compositions provided suitable adjustments are made in the precipitating technique. For example, initial levels of performance are the same for double carbonates of 70/30, 50/50, and 30/70 mole per cent barium and strontium when their particle size is the same. Similarly, triple carbonate compositions of 57/39/4, 44/50/6, 49/45/6,[20] and 56/38/6 weight per cent[41] yield initial levels of emission that are equal when they are precipitated with the same particle size and are given equal treatment of ball milling and spray formulation to yield the same weight, diameter, and density of emission coating. The final choice of a composition ratio, or particle-size range and distribution or both for an emission carbonate is based on standardized practices of manufacturing the carbonate, the sealexing procedures, and life performance of the tubes under the environmental conditions, i.e., materials of construction, processing methods for preparing these materials for use in the tube, and duty cycle requirements in the application of the tube in the circuit.

Under standard conditions of manufacture, composition ratios of double carbonates ranging from 70/30 to 30/70 mole per cent $(BaSr)CO_3$ yield the same particle size. The particle size of the triple carbonate $(BaSrCa)CO_3$ i.e., the calcium component is 4 weight per cent, within the same composition range when precipitated under the same standard conditions is larger than the particle size of the corresponding double carbonate with the same barium content. The crystal structure of the barium-strontium-calcium triple carbonates and the emission performance of the corresponding triple-oxide systems have been investigated by the laboratories at RCA, Harrison. It has been observed that a large sub-central area in the three-component phase-diagram system is capable of yielding higher emission performance than can be obtained from the two-component barium-strontium phase system, i.e., larger particle size for the triple carbonate as a single-phase three-component solid-solution carbonate providing that the calcium component is less than 6 weight per cent. For purposes of such comparison, it is necessary to prepare the carbonates under standard conditions. It is known that the oxide coating derived from a solid solution double-carbonate $(BaSr)CO_3$ system yields higher emission than one from an oxide coating derived from a mechanical mixture of $BaCO_3$ plus $SrCO_3$. For a given heat treatment (temperature and time), a solid solution of the oxide (BaSr)O forms more readily at a lower temperature from the solid solution $(BaSr)CO_3$ than from the mixed single phases of the mechanical mixture. Part of the explanation for the higher emission from the solid-solution $(BaSr)CO_3$ to (BaSr)O can be explained by the larger particle size of the carbonate crystal and the resultant greater porosity (increased conductivity) of the oxide coating. Nonetheless, it is important to take into account the lower temperature of 1325 K versus 1400 K at which solid solution phase formation occurs together with the smaller internal crystallite growth as a contributory factor to the higher conductivity of the oxide coating. These considerations apply equally as well to the three-component single-phase oxide of barium, strontium, and calcium derived from the single-phase three-component solid-solution triple carbonate $(BaSrCa)CO_3$. It is interpreted that the presence of the calcium oxide component serves to complex the oxide phase by reason of the smaller ion radius of the calcium compared to the radii of barium and strontium which would tend to yield a distorted crystallite lattice. The three-component triple oxide tends to possess the smallest parameter crystallite lattice of the three systems (BaO, (BaSr)O, and (BaSrCa)O)[38, 39] and tends to form the single-phase solid-solution oxide with the minimum effects of fusion at the relatively low temperature of 1275 K.

The crystallite formation in the oxide form is determined to some extent by the composition of the carbonate form in that the eutectic melt formation of the barium carbonate component is minimized by the presence of strontium and/or calcium components. Once this eutectic form has been minimized, the interdependence of temperature, time, and carbon dioxide gas equilibrium pressure on the decomposition of the carbonates affects the crystallite growth of the oxide coating. Because the surface energy forces are greater for a fused condition, i.e., the packing density in the lattice structure is greater, the work function of such crystallite systems is higher than that of a system where the melt formation is at a minimum. Measurement of crystallite growth is a measure of the degree of fusion; crystallite growth is more rapid in the melt condition. Investigations[37, 38, 39] indicate that the growth of crystallites in the oxide form is more rapid from the single barium carbonate (eutectic melt condition is pronounced), less rapid from the double carbonate, and least rapid from the triple carbonate. This effect is observed for emission coatings made from single barium carbonate in which the work function is about 1.25 ev as contrasted with the work function of 1.0 ev for the emission coatings derived from the double and triple carbonates. It is noted that at temperatures below 1070 K the reducing agent activity of the cathode base metal is the dominant factor for obtaining high emission from the oxide coating, but at temperatures above 1200 K, heating effects of high temperatures cause rapid growth of crystallites, i.e., fusion, to become the dominant factor for causing low emission performance. Thus, the higher emission that is obtained from coatings derived from the double and triple carbonates over that of the single barium carbonate occurs by reason of the greater conductivity of the oxide coatings derived from the double and triple carbonates. The increased porosity of the oxide coating derived from double and triple carbonates as well as the smaller crystallite growth in the oxide form contributes to this higher conductivity. The higher conductivity of the triple oxide derived from triple carbonates of the same particle size as equivalent compositions of the double carbonate is due to the still smaller growth of crystallites and slightly better conductivity of the oxide coating.

Thus, the use of the coprecipitated triple carbonate of barium, strontium, and calcium contributed several advantages in terms of tube manufacture and tube performance.

1. The particle size of the triple carbonate within the composition range in use as precipitated under standard conditions is larger than the particle size of the corresponding double carbonate of the same barium content.
2. The thermal decomposition characteristics of the triple carbonate are slightly faster and more uniform than those of the corresponding double carbonate by reason of the larger particle size and the earlier decomposition of the calcium carbonate component, provided that the calcium carbonate content does not exceed six weight per cent.
3. The oxide phase resulting from the conversion of the triple carbonate tends to possess the smallest parameter crystallite lattice structure which contributes to the higher conductivity of the coating; the formation of small crystallites in the oxide crystal phase is indicative of minimum fusion or crystal growth, i.e., a greater degree of porosity is retained in the oxide coating.

It is evident in the case of the double carbonate system that a series of solid solutions of $(BaSr)CO_3$ could be made throughout the composition range and that the series of solid solutions could be maintained in the oxide phase (BaSr)O upon conversion of the carbonates to the oxide. However, in the case of the triple carbonate

$(BaSrCa)CO_3$ it is not possible to attain the solid-solution single-phase triple oxide when the calcium carbonate content is in excess of six weight per cent. In the final analysis, since it is desirable to maintain single-phase solid-solution formation of small crystallite size in the oxide form, it is not essential to increase the calcium content beyond the present composition ratio: $(BaSrCa)CO_3$ 57/39/4 weight per cent. The qualities of long life, resistance to poisoning effects (optimum porosity, not maximum porosity) under manufacturing conditions, stability, and minimum quantity-rate of evaporation of barium/barium oxide onto adjoining electrodes are important criteria rather than maximum emission under ideal vacuum tube conditions. Thus, the manufacturing, sealex, and stabilizing schedules would determine the optimum composition and particle size (for a given manufacturing schedule)as the time schedule and heat treatment cycle would determine the single-phase solid-solution oxide form that contributes to high emission performance.

THERMODYNAMIC CONSIDERATIONS OF OXIDE-CATHODE SYSTEMS

In any considerations of the chemical and physical processes that go on during the conversion of the emission carbonates to activated oxide coatings, much thought must be given to a time-temperature-pressure sequence of equilibria reactions. The processes that occur during the decomposition of the carbonates and the subsequent activation of the oxide form influence the properties of the oxide cathode system to a great degree.

Thermodynamic considerations indicate that barium can not be derived from thermal decomposition of the barium oxide nor from the reduction of the barium oxide by the nickel of the cathode base metal. The change in the Gibbs free energy ΔF accompanying a change in the state of matter is expressed as $\Delta F = \Delta H + T(\Delta S)$, where the symbols H and S represent the heat content (enthalpy) and the entropy respectively, and T is the temperature in degrees Kelvin. For a given chemical reaction, if ΔF is negative, the reaction is thermodynamically possible; whereas, if ΔF is positive, the reaction does not take place. Where ΔF is zero, chemical equilibria exists. For purposes of calculation, the relationship of a substance in its standard state (at a temperature of 298 K and at a pressure of 760 millimeters of mercury) between the standard free energy change ΔF^o for the given reaction at the temperature T is $\Delta F_p^o = -2.3\ RT \log K_p$, where ΔF^o is the free energy for the reaction in calories at a temperature T in degrees Kelvin, and R is the molar gas constant equal to 1.987 calories/mole/degree Kelvin. K is the equilibrium constant for the reaction

$$BaCO_3\ (solid) = BaO\ (solid) + CO_2\ (gas)$$

and

$$K = \frac{\text{Conc. BaO} \times \text{Conc. } CO_2}{\text{Conc. } BaCO_3}$$

or, in terms of partial pressures (assuming that solids do not possess vapor pressures), the equilibrium constant $K' = P_{CO_2}$ (partial pressure of CO_2). The free energy change for the transfer of reactants at arbitrary partial pressures to products at arbitrary pressures is given by the equation

$$\Delta F = 2.3\ RT \log K' - 2.3\ RT\ \Sigma n \log p$$

or,

$$\Delta F = \Delta F_p^o + 2.3\ RT\ \Sigma n \log p$$

where $\Sigma n \log p$ represents the algebraic sum of all the n log p terms, those for products being taken as positive and for reactants as negative. The equilibrium constant K' is related to the equilibrium of all partial pressures of all gaseous reactants and products by the rule of mass action and can be applied to the chemical reduction of barium oxide,

$$BaO\ (solid) + Metal\ (solid) = Metal\ oxide\ (solid) + Ba\ (Gas)$$

and, therefore, K' is equal to the partial pressure of barium. Heat and entropy data are usually referred to the reference standard temperature of 298 K; therefore, in order to compute ΔF at a new temperature of 1050 K, it is necessary to obtain data referring to the variation of H and S with temperature in terms of heat capacity data, such that

$$\Delta F_T^o = \Delta H_{T_o} + T\Delta S_{T_o} + f\ (\Delta C_p\ ,\ T)$$

The factor f depends upon the difference in heat capacity ΔC_p between products and reactants and upon temperature T. White[42] and Rittner[43] have indicated the value of such thermodynamic calculations in discussing the rate processes and reactions that occur in the oxide cathode.

Where gaseous products of reactions are formed, the equilibrium pressures impose an upper limit on the rate of reaction, and also on the rate at which these gas products can be removed. It has been indicated that the diffusion process of the reducing agents in the cathode base metal limits the rate of reaction because the reactants can not be brought together fast enough. The following metals can react with barium oxide to yield an equilibrium pressure of barium above barium oxide: Mg, Si, Zr, Al, and W; the following metals are inert in the presence of barium oxide: Ni, Co, Cu, Pb, Au, Pd, and Pt.

The reduction of barium carbonate by metals to yield carbon monoxide is thermodynamically more favorable than the corresponding reduction of barium oxide. For example, the oxidation of silicon or tungsten by barium carbonate will occur preferentially at the lower temperature of 870 K before the reduction of barium oxide by the silicon or tungsten can occur at 1270 K,[44, 45]

at 870 K,	$2BaCO_3$	+	Si	=	Ba_2SiO_4	+	2CO
	barium carbonate		silicon		barium ortho-silicate		carbon mon-oxide
at 1270 K,	4BaO	+	Si	=	Ba_2SiO_4	+	2Ba
	barium oxide		silicon		barium ortho-silicate		barium

Thus, there can be no effective yield of barium released by introducing reducing agents into the carbonate coating. Barium hydroxide and barium nitrate are prone to oxidize nickel and the reducing metals and are not used in emission coatings. Nickel oxide as an interface compound is highly undesirable because it can act as a "sink" for barium as well as being conducive to poor coating adherence. In some instances, a solid solution of $NiO:NiO_2$ can form in the presence of $BaCO_3$ to yield a blue-black barium nickelite, $BaNi_2O_5$.

The decomposition of the carbonate coating is a critical step in the processing schedule. It is during this time that the carbonate may oxidize the surface layers of reducing elements in the base metal and form oxide or interface compounds prematurely along with the release of carbon monoxide. The time for decomposition must be short and is determined in practice by the pumping characteristics of the sealex system and by the temperature of processing. If, however, the time of decomposition is too fast in terms of pumping efficiency, the partial pressure of carbon dioxide will accelerate a melt situation due to the presence of the $2BaCO_3{:}1BaO$ eutectic phase, leading to large crystallite growth and fusion effects. Thus, at the time of decomposition, there are several reaction-rate processes occurring, two of which are competitive for the reducing action of the base metal alloy. For example.

at 870 K, $BaCO_3 + M = BaO + MO + CO\uparrow$

at 1120 K, $BaCO_3 = BaO + CO_2\uparrow$

at 1270 K, $BaO + M = Ba + MO$

below 1300 K, $3BaCO_3 = 2BaCO_3{+}BaO + CO_2\uparrow$ (eutectic)

$= 3BaO + 2CO_2\uparrow$

Thus, under conditions of thermal decomposition of the carbonates, wherein the removal of the carbon dioxide gas is too slow, the undesirable side-reactions leading to the reaction of the carbonate with the reducing element of the cathode base metal at the lower temperature of 870 K, and the concurrent reaction of eutectic melt formation, cause low emission and slumping performance on two counts. The side-reaction with the carbonate uses the surface concentration of reducing elements of the nickel alloy so that the rate of replenishment of barium by the main reaction with barium oxide is slowed down to produce a slumping emission with respect to the normal rate of recovery from gas poisoning actions in the tube after processing and aging. Where the interface compound is formed prematurely, an impedance exists that further lowers the emission performance. The eutectic melt formation tends toward a fusion of the oxide coating, decreasing its porosity, and thus increasing its bulk electrical resistance which lowers the level of emission by reason of the decrease of the cathode to anode voltage.

The activation reaction yielding the barium and associated donor sites in the oxide coating is a "reaction-rate" process defined by the temperature and the following physical-chemical kinetics:

1. Concentration of the effective reducing element in the cathode base metal
2. Lattice structure of the base metal and distribution of grain boundary lines
3. Rate of diffusion of the reducing elements along grain boundary lines in the base metal to the interface surface region[46, 47, 48]
4. Rate of chemical reaction of the reducing element with the barium oxide to form barium and the interface compounds
5. Rate of diffusion of barium through the oxide matrix and rate of diffusion of gaseous barium into the oxide lattice involving equilibrium pressures of barium (solid) to barium (gas) to barium (ion)
6. Rate of electrolytic transport of barium ions to the cathode-metal interface region where it is reduced to the barium
7. Rate of diffusion of barium to the surface of the oxide coating
8. Rate of evaporation of barium/barium oxide from the surface of the coating leaving a layer deficient in barium oxide about 10^{-4} centimeters thick[1]

In order to obtain an appreciable solubility of barium in barium oxide, the vapor pressure of barium must be maintained at a rather high level. It is considered that the barium generated at the interface region is transported through the barium oxide matrix of crystals by a volume-diffusion process termed Knudsen flow, or free molecule diffusion.[43] The Knudsen flow effect postulates that the flow of gas through a long tube (pore space) takes place at a pressure such that the mean free path is much greater than the radius of the tube. Rittner calculates that since the barium can penetrate into the barium oxide particle a maximum distance of 4×10^{-4} centimeters, the oxide coating must be a porous mass containing relatively fine particles in order to account for the rate of evaporation of barium, e.g., the rate of evaporation of barium is four micrograms per square centimeter per hour at a temperature of 1225 K for an oxide coating on a nickel base metal alloy containing 0.12 weight per cent Mg. [43]

As previously stated, a reaction is thermodynamically possible in terms of Gibbs free energy change during any change of a state of matter such that the ΔF is negative. And since the standard free energy ΔF^o is related to the equilibrium pressure through the relation $\Delta F^o = 2.3\ RT \log K'$, where K' varies directly with the partial pressure of the barium gas, the relative activity of the reducing element content of the base metal can be estimated in terms of the partial pressure of barium generated by the molecular reaction of BaO with the element, e.g.,

at 1000 K, $BaO+Mn=Ba+MnO$, the P_{Ba} is 1×10^{-8} mm Hg[43]

at 1000 K, $BaO+C=Ba+CO$, the P_{Ba} is 4×10^{-6} mm Hg[43]

at 1073 K, $BaO+Mg=Ba+MgO$, the P_{Ba} is 7 mm Hg[42]

The thermodynamic lower limit for the equilibrium Ba pressure is that partial pressure of Ba produced by the thermal dissociation of BaO, i.e., 8×10^{-16} mm Hg; the thermodynamic upper limit for the equilibrium Ba pressure is the partial pressure of Ba above a BaO

phase saturated with excess barium which is estimated to be about 1×10^{-2} mm Hg (this limit is somewhat lower than the vapor pressure of pure Ba). Thus, where high pressures are found for active reducing agents, e.g., Mg, it is only because the postulated reaction involving the gaseous Ba phase does not represent the lowest energy state of the system. The calculation does indicate the relative activity of the several reducing elements.

In practice, where the reduction of barium oxide is to be accomplished by a reducing element that is present in very minute concentrations in the nickel base metal alloy, instead of molecular amounts in intimate contact with each other, this type of calculation can not be applied as readily. In order to be able to solve this problem and predict the behavior of the cathode system, it is necessary to have a knowledge of the phase diagram of such dilute-solid-solution alloys, the effective concentration of the reducing element and its rate of diffusion in the alloy; in addition, the free energy of formation of the intermetallic compounds such as MgC_2, Mg_2C_3, SiC, Al_2C_6, WC, Ni_2Mg, Ni_3Si, and Ni_4W must be known before predictions can be calculated.

The reactions between the barium carbonate and the barium oxide forms with the reducing elements in the nickel base metal and with the residual gases in the tube are shown in Table I.

The reactions and equilibria involving carbon monoxide and carbon dioxide are of considerable importance in vacuum tube technology. Around 770 K, carbon (in excess) in an atmosphere of oxygen yields carbon dioxide, but at 1270 K, the reaction of oxygen with excess carbon yields almost pure carbon monoxide:

at 770 K, $C + O_2 = CO_2$

at 1270 K, $2C + O_2 = 2CO$

It follows that the equilibrium between CO_2 and CO at the higher temperature will be toward the formation of carbon monoxide, i.e., $C + CO_2 = 2CO$, in the presence of excess carbon. It is only at temperatures above 2400 K that carbon dioxide is appreciably dissociated to carbon monoxide: $2CO_2 = 2CO + O_2$. Where excess oxygen is present, both carbon and carbon monoxide are oxidized to carbon dioxide: $2CO + O_2 = 2CO_2$. At 870 K, carbon plus water vapor yield carbon monoxide and hydrogen: $C + H_2O = CO + H_2$; and $CO + H_2O = CO_2 + H_2$ occurs at lower temperatures.

Trace reactions can account for methane, alcohol, and even benzene, as follows: $CO + 3H_2 = CH_4 + H_2O$; $CO + 2H_2 = CH_3OH$; $12CO + 3H_2 = C_6H_6 + 6CO_2$.

The reactions of methane with barium oxide at high temperatures are really the reactions of carbon and hydrogen, as the methane decomposes quite readily. Thus, $CH_4 = C + 2H_2$, then $BaO + C = CO + Ba$ and $2BaO + 2H_2 = 2H_2O + 2Ba$. The reaction of methane with barium oxide gives rise to a high concentration of barium but results in low emission, nonetheless, because the carbon deposits onto the coating to yield a dark surface with a resultant low operating temperature, i.e., with black body radiation.

DECOMPOSITION OF THE EMISSION COATING

The decomposition of the carbonate coating is a critical step in the processing of an oxide cathode system. In the decomposition process, the carbon dioxide gas is in equilibrium with the barium carbonate and barium oxide phases at definite dissociation pressures with temperature. Because the dissociation pressure for barium carbonate is the lowest of the three component carbonates, it is the determining factor in the selection of the temperature for decomposition. It is during the time that the carbonate is being converted to the oxide that the competitive reactions of the reducing elements with barium carbonate and barium oxide make it mandatory to convert to the oxide phase as quickly as possible in order not to deplete the surface concentration of reducing elements in the base metal before they can react with the barium oxide to yield the active barium. The time for decomposition should be short and is determined in practice by the pumping characteristics of the system and by the temperature. However, when the time for decomposition is too short relative to the pumping efficiency of the system at the surface of the coating, the partial pressure build-up of carbon dioxide gas will repress the decomposition of the carbonates long enough to allow for the direct oxidation reaction of the carbonates with the reducing elements, and also, will create the eutectic melt phase that produces large crystallite growth and fusion effects. If silicon

Table I

Chemical Reactions of $BaCO_3$ — BaO with Some Reducing Elements

Reducing Element	Temperature	
	870 K	1270 K
carbon	$BaCO_3 + C = BaO + 2CO$	$BaO + C = Ba + CO$
magnesium	$BaCO_3 + Mg = BaO + MgO + CO$	$BaO + Mg = Ba + MgO$
manganese	$BaCO_3 + Mn = BaO + MnO + CO$	$BaO + Mn = Ba + MnO$
silicon	$2BaCO_3 + Si = Ba_2SiO_4 + 2CO$	$4BaO + Si = Ba_2SiO_4 + 2Ba$
tungsten	$3BaCO_3 + W = Ba_3WO_6 + 3CO$	$6BaO + W = Ba_3WO_6 + 3Ba$
carbon monoxide	- - - - - - - - - - - - - - - - - -	$BaO + CO = CO_2 + Ba$
hydrogen	$BaCO_3 + H_2 = BaO + CO + H_2O$	$BaO + H_2 = H_2O + Ba$

or tungsten are involved as reducing agents, premature formation of the interface impedance may occur due to the formation of the barium orthosilicate or barium tungstate to yield the consequent lower emission performance. Thus, the sealex operation becomes a complex sequence of events related by the factors of time, temperature, and pressure during the decomposition of the carbonate coating, as shown in Fig. 12,[49] and the diffusion-rate-reaction process of the reducing-element content of the cathode base metal together with the interrelated gas and vapor reactions from adjoining tube elements.

The partial pressure of carbon dioxide gas in the tube during decomposition of the coating on the sealex is determined not only by the size of the cathode and the weight, density, and particle size of the carbonate coating, but also by the rate of exhaust, and the pumping efficiency. The factor of pumping efficiency is influenced by the time, temperature of decomposition, volume of the tube, gas content of the tube, diameter of the tubulation, and the bore size of the plumbing. The vacuum during the decomposition of the carbonates is about 10^{-3} millimeters of mercury on sealex so that a breakdown temperature of 1273 K would be sufficient, provided that the time for decomposition were of the order of tens of seconds. However, production sealex speeds of 1000 units per hour allow only about 2.4 seconds per position and 1.2 seconds for indexing to the next position. Thus, temperatures for initial decomposition of the carbonate coating actually range about 1473 K in order to bring about rapid breakdown of the carbonate coating in the first few positions. X-ray analysis shows that the single-phase solid-solution oxide form occurs about 1350 K; correlating data indicate that the high emission is obtainable from such a small-crystallite solid-solution oxide matrix. Therefore, it is desirable to attain a high temperature for some short period of time in order to obtain good emission performance. This result can be achieved by heating the cathode to a high temperature provided certain side-effects do not overshadow the desired goal of a good activated oxide cathode in a tube possessing good operating characteristics. At high temperatures, the following side-effects must be taken into consideration:

1. The more volatile barium/barium oxide can evaporate preferentially from the triple-oxide matrix to an excessive degree, as shown in Fig. 13[50] and Fig. 14[51] and contaminate adjoining grid electrodes and thus tend to introduce effects of primary grid emission, contact potential shifts, and even secondary emission from other electrodes.
2. The crystallite structure of the oxide coating can grow too large and fuse to reduce the conductivity of the coating, and thus, lower the level of emission performance.
3. The evaporation of the more volatile chemical reducing agents alloyed with the cathode nickel, as shown in Fig. 15,[52] e.g., Mg, Mn, can cause leakage paths across adjoining surfaces which would tend to produce feedback, rf noise, and lower transconductance.
4. The nickel cathode base metal (melting point 1450 C) can be softened, i.e., lose hot bend strength, to create bowing of the cathode and misalignment.
5. The tungsten heaters may become embrittled by excessive crystallization and burn out; the insulation

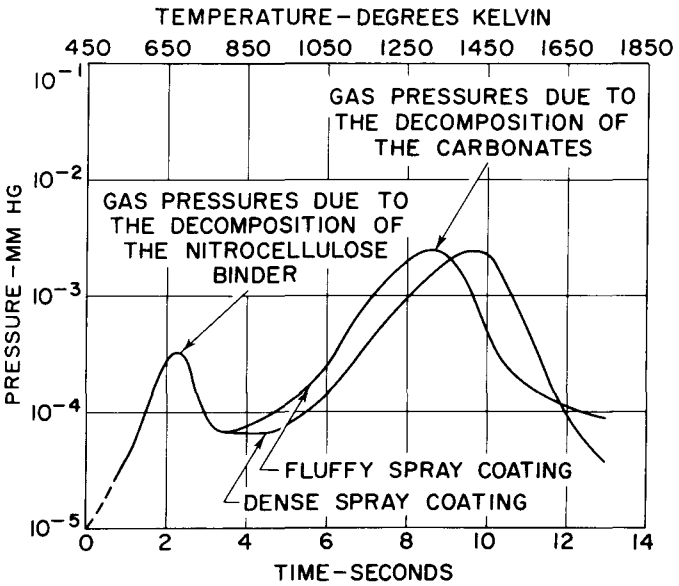

TEMPERATURE - DEGREES KELVIN
450 650 850 1050 1250 1450 1650 1850
PRESSURE - MM HG
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
GAS PRESSURES DUE TO THE DECOMPOSITION OF THE CARBONATES
GAS PRESSURES DUE TO THE DECOMPOSITION OF THE NITROCELLULOSE BINDER
FLUFFY SPRAY COATING
DENSE SPRAY COATING
0 2 4 6 8 10 12 14
TIME - SECONDS

Figure 12. Pressure-Time-Temperature Interdependence on Decomposition of the Emission Carbonates[49]

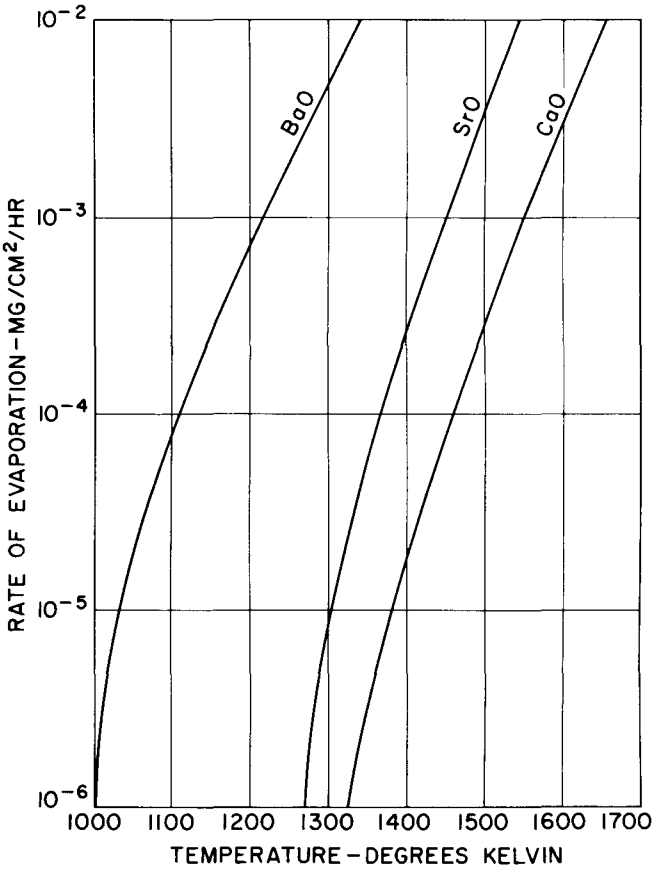


Figure 13. Rate of Evaporation of Alkaline-Earth Oxides[50]

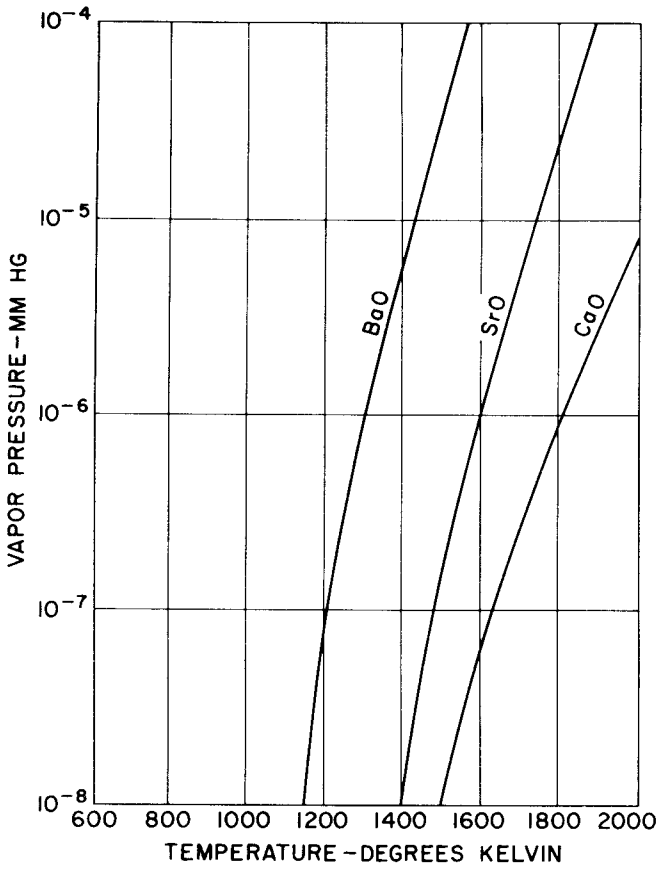


Figure 14. Vapor Pressure of Alkaline-Earth Oxides[51]

coating can become impaired and create heater-cathode leakage effects.

Hence, in practice, there will be an optimum temperature for decomposition; this temperature determines the minimum time for the decomposition of the carbonate coating, provided that the partial pressure of carbon dioxide gas is not excessive.

It can be observed from the curves of the dissociation pressures of the alkaline-earth carbonates, as shown in Fig. 16, that barium carbonate is very difficult to decompose because of the low equilibrium pressure of carbon dioxide that exists at the lower temperatures. The reaction can be driven to completion only by rapid removal of the gas under vacuum pumping conditions. The decomposition of the carbonate coating is primarily a problem of gas evolution-gas removal at the surface of the coating at a given temperature regardless of the vacuum pressure in the tube as a unit that is measured at some distance from the cathode surface. If the condition of temperature and pressure of carbon dioxide gas at the surface of the coating places the system to the right of the equilibrium pressure curves, the carbonates will decompose without giving evidence of pitting or fusion, i.e., the eutectic low melting point phase of the barium carbonate will not have been formed to an appreciable extent. Should the temperature-pressure equilibrium that exists at the surface of the coating place the system to the left of the equilibrium curve, the barium carbonate component will not decompose quickly enough to avoid formation of the eutectic phase. The entire sequence of events is transient in time (fractions of seconds) such that the pumping efficiency, tubulation bore size, gas content and surface areas of the anodes, and the amount and density of carbonate coating are in critical balance with sealex speed, heater current, and rf induction heating.

After decomposition of the carbonate coating has been accomplished during the early stages of sealexing, the envelope and electrode parts of the tube are further degassed in successive stages of induction heating while the cathode is kept hot by means of the lighted heater. When the cathode coating is in the oxide form, it is vulnerable to the poisoning actions of gases and water vapor that would tend to change it back to the hydroxide or carbonate. After degassing, the tube is "gettered" (the getter is flashed) to absorb residual gases, and then is sealed off. Usually, there has been sufficient chemical reducing activity to create the necessary barium phases in the oxide coating by the reaction of the barium oxide with the reducing agent. A high emission current can be obtained immediately after decomposition of the coating; however, it is often necessary to activate the oxide coating still further after the tube has been sealed off because the amount of barium and associated active donor sites are depleted by reaction with the gases coming from the adjoining electrode surfaces upon electron bombardment during the passage of current in the tube.

Thus, after gettering and sealing off, the tube is usually heated to a high temperature by the use of one and one-half times the rated heater voltage in order to create additional active barium centers by chemical reducing action of the cathode base metal alloy on the oxide coating. Then the residual gases must be removed from the adjoining tube electrodes. This is accomplished by applying positive voltages to grid and anode and drawing current. The electrodes are bombarded by electrons and heated so that their gases are driven off to be absorbed by the getter and by the activated cathode coating. During such treatment, the emission current passes through a minimum and then increases to a higher value as the gases are absorbed by the getter and cathode coating. During this "aging" schedule, the rates of diffusion of reducing elements in the base metals are stabilized and active barium centers in the coating are well distributed by electrolysis and chemical reducing activity. Details of an activation schedule will vary with the tube type and kind of material used in construction. The size and configuration of the electrodes, the properties of the materials, degree of preparation by washing and hydrogen firing, type and position of the getter, cathode base metal alloy (active or passive), coating weight and density, and the efficiency of the original sealex degassing and pumping will determine the activation schedule.

AMMONIUM-PRECIPITATED CARBONATES

The porosity of the oxide coating, i.e., the degree of fusion and crystallite growth, is adversely affected by low-melting impurity salts that may be present in

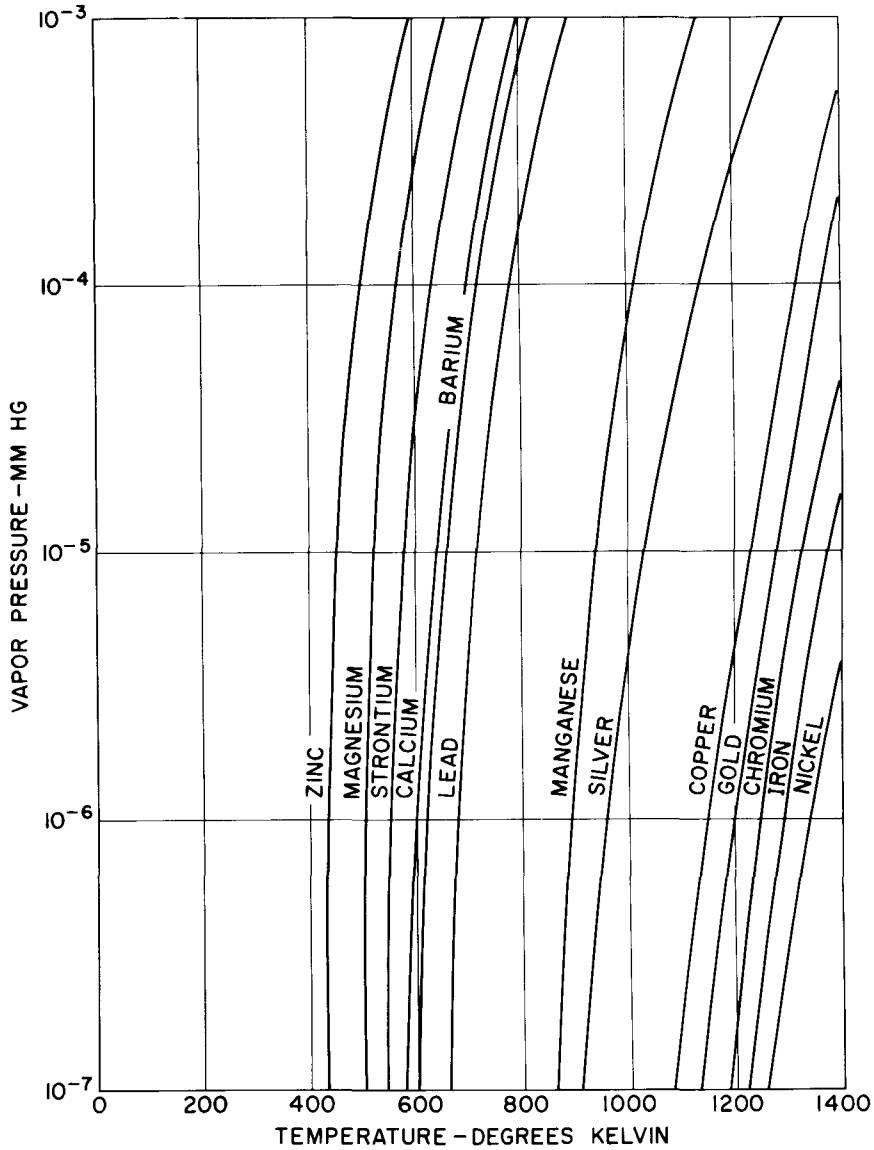


Figure 15. Vapor-Pressure Curves for Some Elements[52]

the emission-carbonate form. As far as can be determined, the presence of low-melting solium-salt compounds as an absorbed layer at the crystal surfaces of the carbonate causes excessive sintering action between adjoining crystals during the low temperature range (500 to 750 K) of decomposition. Later in the time sequence (measured in milliseconds), the internal content of sodium salts that is released upon complete decomposition of the carbonates contributes to the low-melting condition of the barium carbonate eutectic phase at the higher temperature range of 1000 to 1270 K. Furthermore, it is at this temperature range in the decomposition process that the more volatile sodium/sodium oxide molecules are deposited upon adjoining electrode surfaces. Thus, this volatile sodium contaminant contributes to excessive leakage and RF noise effects beyond the normal level that is ordinarily contributed by deposits of barium/barium oxide.

The ammonium-precipitated type of emission carbonate is purer than the emission carbonate prepared by use of sodium carbonate reagents because any ammonium salts left in the emission carbonate, after precipitation and washing, vaporize completely as gases. Thus, the oxide coating is purer than the parent ammonium precipitated emission carbonate. Because the oxide coating derived from ammonium-precipitated carbonates contains no contaminant with a low melting point (the eutectic phase is the sole inherent melt condition), it does not sinter and contract in dimensions to the same extent as oxide coatings containing sodium salt impurities. The spacing between cathode and grid does not change as drastically because the sintering action is at a minimum so that transconductance decreases less on life.

In general, the contamination of grid wire surfaces is the result of barium/barium oxide evaporating from the oxide coating. This normal form of contaminant is more stable than where the more volatile sodium/sodium oxide forms are present so that the tube perfor-

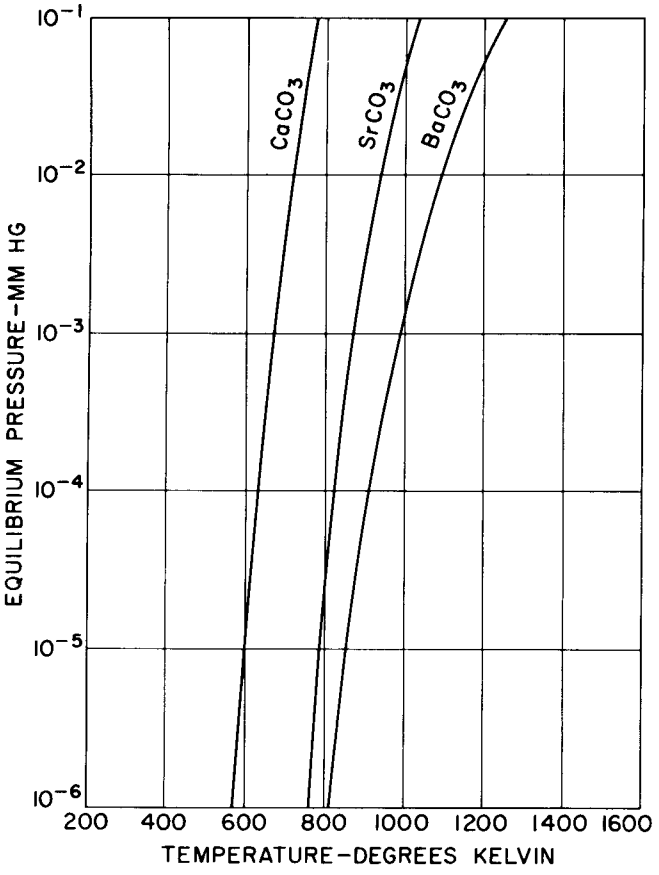


Figure 16. Equilibrium Pressures of the Alkaline-Earth Carbonates

mance slumps less over the life period as the contact potential difference between cathode and grid is more constant. The more stable type of contaminant of barium/barium oxide, ordinarily arrived at by the use of slightly higher controlled hot shot schedules during aging, can not be usefully obtained when using carbonates having sodium-salt impurity content because of the excessive leakage effects induced by the presence of the more volatile sodium/sodium oxide contaminant on grid and mica surfaces.

Compared to the oxide coating derived from sodium-precipitated emission carbonates, the oxide coating derived from ammonium-precipitated emission carbonates is more porous. As such, the latter form reacts more quickly to the poisoning actions of gases in the tube. However, when the tube has been properly degassed, the very porosity of the oxide coating derived from ammonium-precipitated carbonates offers higher conductivity across the coating by reason of the electron-pore gas effect. An improved level of emission performance can be obtained, with more stable contact potentials, minimum screen-grid emission, and minimum decrease of transconductance on life. The important corollary to gas poisoning actions on the oxide coating is the need for proper parts, properly preprocessed. In order to manufacture tubes of good quality, and with minimum shrinkage due to poisoning effects, the proper selection of materials of construction for anodes, i.e., low gas content together with good heat dissipation capability, or, better parts processing in terms of hydrogen firing and sealex degassing would be required.

LIFE ASPECTS OF THE OXIDE-COATED CATHODE

The oxide cathode is usually spoken of as an electrode in the vacuum tube. Actually, it is a complex system of interacting phase changes between the nickel base metal and the oxide coating derived from alkaline-earth carbonates. Specific alloys of nickel containing controlled trace amounts of several metallic agents of magnesium, silicon, manganese, tungsten, and even carbon are used to obtain the desired degree of activity from the cathode system. These reducing elements react with the barium oxide to yield barium and associated active donor sites in the oxide lattice at a replenishing rate during the life of the system. The rate of diffusion of these reducing elements is the controlling factor in determining the activity and recuperative power of the cathode during the life of the tube.

The emission carbonates from which the oxide coating is derived has been developed along a range of composition ratios with respect to barium, strontium, and calcium. These carbonates are carefully controlled with respect to size, shape and purity. Moderate bulk density, needleform carbonates are used for most of the applications where high level emission performance is necessary; high density, spherulite forms of carbonates are used for emission coatings where the requirements demand smooth, dense coatings for close-space, high-voltage gradient rectifier service.

After being applied to the cathode by suitable means of spray coating or electrophoretic techniques, the emission coatings are held in place by the binder of nitrocellulose or methyl methacrylate (Lucite) that was present in the carbonate coating suspension. After the coated cathodes have been assembled and mounted in the tubes, the carbonates are converted to the oxide by a combination of heating and vacuum pumping to form an adherent coating of oxide crystals on the cathode base metal; carbon dioxide and carbon monoxide gases, along with other tube gases, are removed by vacuum pumping. Nitrocellulose is oxidized to several forms of nitrous oxides, water vapor, carbon dioxide, and light organic fractions of ethylene at 200 C; where methyl methacrylate is used as the binder, it sublimes as the monomer gas phase at 300 C. A carefully controlled schedule of high-temperature filament heating of the cathode and rf-induction heating of the metallic parts is performed on the sealex. Further high temperature treatment is performed on the cathode, and actual current passage is accomplished during the aging-stabilization step to bring the cathode system to a reasonably high level of emission activity.

The mechanism of electron transfer through the oxide-coated cathode has not been completely detailed in all its phases. There exists a basic semiconductor mechanism for electron transfer coupled with an electron pore gas path of transfer which improves the conductivity of the system. The equilibria phases of barium (solid) to barium (gas) to barium (ion) are involved such

that a large excess of barium vapor phase is required to maintain a threshold content of active donor sites in the oxide lattice. Changes occur in the surface layers of the oxide matrix due to the preferential rate of evaporation of barium/barium oxide from the external surfaces. Strontium and calcium oxides which have lower vapor pressures than barium oxide at the operating temperatures of the cathode system remain after the outer surface is free of barium oxide. Electron-diffraction techniques confirm the fact that for equimolar barium-strontium oxide, the surface layer is almost pure strontium oxide. Eisenstein was able to calculate from X-ray scattering patterns that the surface composition differed from the bulk composition of the oxide.[1] In 115 hours, the surface layers containing less than 10 mole per cent of barium oxide were about 10^{-5} centimeters thick; at 1205 hours, the layer that was deficient in barium oxide was about 3×10^{-4} centimeters thick. In this instance, the cathode coating that was being observed was heated but no current was drawn. Because the physical nature of the external surface of the oxide matrix is independent of the bulk oxide composition, electron affinity for such surfaces would appear to be the same regardless of the bulk composition. In practice, a range of bulk compositions exists that yield the same emission performance when due regard is given to the carbonate particle size and the resultant porosity-conductivity of the oxide coating.

When the emission obtained from an oxide coating is decreased by gases or other foreign material, the term "poisoning" is used to denote a series of physical-chemical reactions that tend to deplete the amount of active donor centers in the oxide lattice as well as to bring about physical changes in the coating that causes its bulk resistance to increase. The poisoning of the oxide coating is most likely to occur at low operating temperatures of the cathode wherein the forward equilibrium reactions creating the active barium centers are slowed down to the extent that the reverse physical-chemical phenomena of depleting the active barium donor content override it.

These "poison" gases or vapors may be given off by the nickel base metal, by the micas, by the heaters, by the anodes, and by the glass envelopes.[53,54] Water vapor combines with barium oxide or barium to form barium hydroxide which, as a low-melting compound, induces excessive sintering and fusion of the cathode coating.[55] Chlorides and fluorides form low-melting compounds with barium, too. Fused coatings have a minimum porosity such that the conductivity of the coating is decreased, i.e., bulk electrical resistance is increased. Gas poisoning due to carbon dioxide, carbon monoxide, and oxygen are, in effect, the result of harmful equilibrium pressures of these gases that tend to extend the low-melting actions of the eutectic phase of barium carbonate. Where the barium phases are depleted due to the interaction with these gases, the cathode system can bring about a recovery from the effect by the replenishment of the barium by means of the reducing activity of the agents in the nickel alloy. Where the malfunction has been brought about by sintering of the coating, no recovery can be obtained; further processing by activation-aging schedules only worsens the melt situation and increases the bulk resistance of the coating.

The pulse operation of the oxide-cathode system can give initial high currents of 50 to 100 amperes per square centimeter during one or two microsecond intervals provided that the duty cycle (repetition rate) is not excessive. The initial high level of current may decay swiftly due to impedance effects of four basic phenomena:

1. Surface layer "skin resistance" at the oxide-coating-to-vacuum interface. This impedance is thought to be caused by a depletion of active emission centers by reason of a depletion of barium ions which are rapidly transported away from the surface by the large electrolysis effects of high current densities. The removal of positively charged barium ions from the surface layers would leave a relatively high negative field at this surface for the transient one- or two-microsecond interval. Correspondingly, the excessive loss of electrons from the surface during the high-current passage exerts a cooling effect greater than the Joule heating effect in the bulk of the oxide matrix such that the surface layer is at a lower temperature and therefore maintains a high impedance.[27]
2. Interface resistance due to the formation of barium orthosilicate by the chemical reaction of silicon with barium oxide. This type of interface compound, where it exists, offers impedances of 50 to 200 ohms which introduces a large voltage drop — cathode to anode—during large current passage.[56,57]
3. Insufficient plate-dissipation capability causes gases to be evolved upon electron bombardment of the anode. These gases can react with and deplete the active donor centers of the oxide coating to reduce the conductivity of the coating so that the emission level of performance will be decreased.
4. Blocking action due to the formation of the equivalence of an electron space charge effect in the pore spaces. The large micro-second surge of pore-gas electrons, including high ratio secondary electrons, actually builds up rapidly to exert a space-charge limiting effect within the pore volume to throttle itself back and thus create the decay effect. Thus, rapid cascading avalanche type of secondary electron build-up may be applicable in the instance of the oxide cathode under pulsed conditions.[6]

Normal tube life expectancy is temperature limited. The lower limit of temperature is determined by the level of emission desired and the reserve capacity required to resist poisoning actions; the effects of poisoning increase with decreasing temperature of cathode operation. The upper limit of temperature is limited by several factors, i.e., evaporation of barium/barium oxide, sublimation of base metal reducing elements, sintering of the coating, and interface formation and growth. It has been calculated that 50 per cent of the barium oxide content of the coating would evaporate within 1500 hours at 1200 K, whereas, only 83 per cent of the barium oxide would evaporate within 1,000,000 hours at 1000 K.[51]

The requirements of reliability, stability, and long life from vacuum tubes involve two major areas of tube technology, i.e., the tube-manufacturing operation and the application of the product in electronic circuitry.

The Manufacturing Operation

From the manufacturing point of view, the problems involved include mechanical failure and electrical failure induced by physical-chemical phenomena. Gross mechanical errors of misalignment, short circuits, vibration hazards and glass breakage, and mechanical failures due to burn-out of heaters, bowing of cathodes and thermal shock can all be minimized by correct design and by the use of selected materials of construction. Electrical failure, on the other hand, that is induced by gaseous products of evaporation and chemical interaction across electrode surfaces is mainly involved in the manufacturing operation of preparation of parts, of maintenance of efficiency and cleanliness of sealex equipment, of the mounting operation, and of observance of the limitations of the materials of construction, e.g., in the use of copper-clad side rods for cooling effect on the grid laterals; one must be aware of the relatively high vapor pressure of copper and its evaporation under too high a setting of the rf induction-heating equipment used for degassing the metal parts.

The physical chemistry of the oxide cathode has been discussed in some detail. The recognition and understanding of the reactions and interactions are important in order to establish a suitable manufacturing process. Controlled purity of the cathode nickel alloy and of the emission carbonates is necessary in order to obtain reproducible results and maintain a manufacturing schedule. The cathode nickel alloy, in terms of its reducing element content and diffusion rate processes, controls initial-life as well as long-life characteristics of the cathode system by its ability to contribute a replenishment rate of supply of active barium emission centers. The oxide coating, in terms of matrix resistance, space geometry, and porosity of structure, effectively controls the level of emission and transconductance values.

These factors are interdependent and are acted upon by the environmental gas and vapor phases in the tube at the time of decomposition of the coating. The geometry of the tube and the sealex speed required for optimum production define the temperature limits for the decomposition of the coating consistent with pumping efficiency and gas removal; these limits, in turn, determine the requirements of weight, outside diameter, and density of the emission carbonate coating. The equilibrium-gas pressures at the transient port and index times on the sealex influence the direction and relative speed of chemical reactions necessary for good activation and for proper porosity structure within the oxide coating. The gas content of the other electrodes in the tube will influence the partial pressures of carbon dioxide and carbon monoxide gases, which, in turn, will affect the decomposition characteristics of the carbonate coating. The optimum conditions for the proper sequential processing of tubes in the manufacturing operation (provided all component parts are properly cleaned and prepared) should consider the following:

1. Glass envelopes should be baked out just prior to use (preferably during the process) to remove much of the air and water vapor that are absorbed in the glass. The release of such gases and water vapor during tube processing and life poisons the activated oxide coating by reacting with and depleting the active barium centers as well as accelerating sintering of the coating.
2. Sealing-in flames should be adjusted, and the mechanical alignment of the cathode within the tube should be arranged so that the coating is not subjected to temperatures greater than 200 C. The nitrocellulose binder volatilizes while being oxidized at 200 C leaving the residual carbonate coating poorly adherent to the base prior to decomposition. The result is an incomplete bonding of the coating to the base metal such that peeling of coating can occur initially as well as later on in life. Moreover, when such a poor interface bond exists, proper activation of the coating can not be achieved since good paths for the diffusion of the reducing elements from the base metal do not exist.
3. The "breakdown positions" — i.e., normally, No. 5 and No. 6 in a 16-head sealex operation — should be adjusted for heater lighting and rf-induction heating such that there is a relatively rapid decomposition of the carbonate coating to the oxide form. This cycle is admittedly the result of a balance of many factors among which are the size, weight, and density of the carbonate coating, the total gas content of the adjoining metal electrodes, the warm-up time of the heaters and the orientation of the rf coil for optimum induction heating. The controlling mechanical feature (all other things being equal) is the bore size of the tubulation which determines the efficient removal of the gases by the pumping system. The complex state of affairs that must be balanced is the effect of the relative carbon dioxide gas pressure at the time of breakdown. Where the time-temperature-pressure ratio results in a high relative pressure of the gas, the eutectic melt phase of the carbonate is prolonged to the extent that the resultant oxide coating has a lesser porosity and in extreme cases may be fused. The relatively high partial pressure of carbon dioxide gas at the time of decomposition of the carbonate coating accelerates the oxidation of the reducing element content at the surface of the base metal without creating active barium centers in the coating. The net effect is premature heavy interface formation, low emission, and early life slumping which will require a longer and harder aging schedule to bring the cathode system and tube to a good level of emission performance.
4. The position of the getter in the tube and the position setting of the getter flash coil are of critical importance in creating a good tube. The position of the getter should be such as to eliminate a line of sight path to the cathode in order to prevent poisoning of the coating by the deposition of metallic elements from the getter channel or the getter compound itself. Deposits of getter material onto the insulating surfaces of micas and electrode supports create leakage effects. The use of metal or mica shields can minimize such effects. The getter flashing coil on the sealex should be positioned such that no extensive heating of the anode metal surfaces occur to release gases onto the oxide coating. In this position in the sealex sequence, the oxide coating is not being kept hot while heater lighting is not being applied so that the coating is vulnerable to the poisoning effects of

the oxygen and water vapor that are being released from the glass walls and the metal parts.

The production of a uniform highly reliable tube product is dependent upon parts preparation and processing. Uniformity of cathode-base-metal structure and effective reducing element content (vacuum-melt alloys); uniformity in carbonate-coating weight, density, and outside diameter; well degassed or essentially gas-free parts; and stable line voltages to maintain uniform heater-lighting and rf-induction heating schedules would contribute to this uniformity and reliability. It was indicated earlier that activation can be achieved by chemical reducing activity to create barium centers without the passage of current. Many tube types come off the sealex in an active condition solely by means of thermal action on chemical activity. The use of the aging schedule, after sealex processing, is mainly for the purpose of stabilization and gas clean-up of metal parts by electron bombardment during the passage of current. Correct settings will yield uniform and reliable results: meters should be calibrated; panel wiring, resistors, and ballast lamps should be operative and at rated dissipation wattage. Correct wattages and time sequences will insure reproducible results. In general, hot shot treatments (high heater wattage application without current passage to activate the coating by thermal chemical means) should be at relatively lower filament voltages and longer time sequences rather than at higher filament voltages and shorter time sequences in order to avoid fusion of the coating, excessive rates of evaporation and sublimation from the cathode system that causes leakage, and bowed cathodes.

Application End Use

From the operational point of view, the problems involved in insuring tube reliability and long life require consideration of the limitations of materials of construction based on the recognition of the dynamic equilibrium of the physical-chemical processes involved in creating electron emission. Tube design and tube application should take into consideration the following basic ideas.

1. Cathode systems should be designed to operate at the relatively low temperature of 1000 K for long life. The cathode system should be operated within ± 5 per cent of this rated value in order to obtain optimum life performance.

At operating temperatures above 1050 K, the evaporation of barium/barium oxide is accelerated so that cathode life expectancy is decreased. Deposition of such barium/barium oxide onto adjoining insulating support surfaces creates leakage paths and results in variable frequency response. Deposits of such material onto grid Lateral wires give rise to primary grid emission, reverse grid current which changes the bias, rf noise, shifts in contact potential, and even blocking-action effects.

At operating temperatures below 950 K, the emission level will decrease due to the slower rate of diffusion of the reducing element content of the base metal required to maintain the replenishment rate of active barium centers needed to compensate for the poisoning action of the normal tube vapor phase. Sparking and arc-over may result due to the high voltage gradient with respect to local hot spots. Cold starting, i.e., application of anode voltage before the cathode has reached operating temperature, should be avoided where possible because positive-ion bombardment will damage the cathode system to the extent of rupturing the oxide-coated surface. Normally, the effects of positive-ion bombardment are minimized because the space cloud of electrons (surrounding the heated cathode system) cushions and neutralizes the velocity of these positive ions.

2. Envelope temperatures should be maintained at reasonably low levels. Glass and metal surfaces adsorb and absorb quite an amount of water vapor and gases; such surfaces, upon being warmed, release these materials as vapors which poison the cathode in the extreme case, and cause slumping performance, grid current flow, and back-bombardment effects in general. For similar reasons, the plate dissipation capability should be adequate to minimize excessive gas release with the wattage required to be handled. The designed factor of safety may be nullified if the ambient temperature of the tube is elevated above that expected in normal tube application. It should be noted that a deposit of getter-flash material on the sidewalls of the tube, where it opposes the radiation characteristics of the anode, can actually reflect 80 per cent of the heat that would have passed through the glass. Therefore, tube and application design should make provision for proper ventilation in order to maintain a low ambient temperature. Where the application requirements demand a higher rating, either a redesign of the tube structure is indicated, or a change in the materials of construction within the same tube design, e.g., substitution of a copper-core, iron-clad, aluminum-clad anode metal for a nickel-clad-iron metal.

3. The maximum tube ratings are limits below which the tube must be operated in order to obtain good performance. In applying ratings of a tube, the equipment designer must take into account the effects of longer duty cycles or higher operating frequencies. Longer duty cycles can impose heavy current demands on the cathode coatings which can cause it to become heated excessively and cause it to sinter. High frequencies can increase tube losses which decrease the output of the tube and reduce its efficiency.

4. Tube characteristics do change on life due to the chemical-physical nature of electron emission that is a function of coating porosity, i.e., characteristics can be changed by sintering of the coating which lowers the conductivity or increases the bulk electrical resistance and also increases the spacing between cathode and grid,[58] by decrease in the effective reducing-element content of the base metal, by decrease in the effectiveness of the getter, and by electrolysis effects on glass which release gas. Thus, circuits should be designed to operate over a rather wide range of tube characteristics. A range of ± 20 per cent of published values is a commonly used criterion. Acceptable life performance can be extended in many instances if a wider range of plate current and transconductance values can be tolerated. This statement is not intended to imply a looseness in rating which

permits nonuniformity of characteristics; what is intended is that the circuit application of a specific tube be evaluated on the merits of the tube design rather than on the requirements of the circuit. If the circuit requirements are mandatory, a tube of the proper design should be used that can meet the requirements. Some application problems originate from lack of consideration of the limitations of the tube rather than from the defects of its design or manufacture.

5. Formation of interface impedance formerly interfered with standby and cutoff type of operation. The use of high silicon (0.08 per cent) nickel alloy as a cathode base metal would normally create the interface compound, barium orthosilicate, to an excessive extent. The problem is not as extensive today as formerly because most present-day cathodes are of the low (0.02 per cent) silicon type. However, where this problem can still cause malfunctioning because the circuit is operated most of the time well-biased below cutoff, reduction of the heater voltage during periods of cutoff or standby operation will minimize the growth of the interface compound.

6. The problem of grid contamination by evaporation of barium/barium oxide from the cathode system is essentially due to the intensity of the processing technique (high temperature) rather than the cumulative effect during life.[59] The problems of contact potential, grid emission, and developed bias readings are discussed more thoroughly under their specific chapters.[60] However, where negative grid currents may be due to primary grid emission, ionization of residual gases, or leakage currents across micas, high values of grid resistors increase the susceptibility of the circuit to develop a runaway condition. In some instances, the proper choice of a grid resistor of slightly lower value (if it did not change the circuit condition) made the difference between a good tube and a bad one assuming no excessively contaminated grid surface or gas content in the tube.

7. The problem of noise and hum has become more severe in today's high-frequency tubes having closely spaced elements. Noise may be attributed to each of the following: emission coating and gas ionization, as well as electromagnetic, contaminating, and mechanical sources.

a. Shot effect is the noise derived from random bursts of electrons propagated under conditions of temperature-limited emission. The shot effect arises from the discreteness of the electron; the effect has the same intensity at all frequencies.[61] A reduced shot effect is experienced under space-charge limited conditions; the noise is due primarily to velocity variations of the electrons in the potential minimum region.

b. Flicker effect[62, 63] is propagated by variations in the activity of the emitting surface of the oxide coating. This effect is essentially a low-frequency variation rather small for space-charge limited conditions but large for temperature-limited conditions. The flicker noise effect is induced by the passage of varying velocity electrons across the charged layer set up on the surface of the oxide coating. Pore gas electrons have different escape velocities induced by them by the effects of electrostatic gradients through the bulk of the oxide matrix. The surface dipole layer comes into existence due to the electrolytic transport of active donor centers back to the cathode and the escape of electrons into the vacuum space during the passage of current. The consequence of the passage of varying electron current across a charged layer is to induce an amplitude-modulated signal of rather low frequency variation. This effect can be minimized to some extent by the use of high-density coatings which result in a reduced porosity of the coating, thus lowering the velocity of these electrons. The use of an oxide coating with decreased porosity reduces the level of emission as rated by the d.c. diode emission test, but there is still sufficient current for normal space-charge limited conditions.

c. Gas ionization effects are induced by variable positive-ion gas currents in the grid circuit. The effects of this noise are not appreciable unless the positive-ion current is greater than 0.02 microampere and its frequency is below 10 megacycles.

d. Electromagnetic induced noise at frequencies above 30 megacycles result from induced voltage variation in the grid circuit due to stray electromagnetic fields; these grid-voltage changes cause the plate current to vary.

e. Reverse grid current noise is induced by negative grid currents, created by positive-ion bombardment or primary grid emission, across composition resistors in the grid-leak circuit.

f. Leakage noise arises from the variable leakage currents along mica insulators and glass due to the contaminants evaporated from the cathode base metal and the oxide coating, e.g., Mg or Mn from the nickel cathode, Ba/BaO from the oxide coating. The phenomenon is closely related to the manufacturing operation, the materials of construction, and the circuit application. Excessive heater wattage during sealex and aging schedules will accelerate the volatile elements (Mg, Mn) out of the base metal along with amounts of Ba/BaO from the coating. The choice of base metal with respect to a low content of Mg or Mn will determine whether a given heat treatment for activation is or is not excessive with respect to the metal. The density of the oxide coating will influence the quantity-rate of evaporation of Ba/BaO. The choice of mica coating or glass surface can determine whether leakage can become a factor as a consequence of processing temperatures. The end application wherein a 12-volt tube is required to pass a 16-volt filament life test, will determine the original choice of material to an even lower content of volatile-reducing-agent content in the nickel base metal, or to a more dense oxide coating.

g. Mechanically induced noise results from heater hum from the use of non-inductive heater windings; this noise may be cancelled out by proper design and choice of winding, i.e., the use of double-helical-wound heater types will cancel out 60-cycle variations in magnetic fields. Noise currents due to heater-cathode leakage effects resulting from alkali impurities in the heater coating can be minimized by the use of pure heater coating material or improved heat transfer characteristics at lower

operating temperatures for the same wattage input.

h. Poor mica insulation can allow more than normal leakage currents to flow because its surface is more conductive, or the condition of the mica coating did not provide sufficient roughness to interrupt the continuity of the leakage paths. This effect can be minimized to some extent by the proper choice of mica coating or texture of coating applied to the mica surface.

i. Vibration can cause noise currents to flow. When the vibration is due to loose fit of the cathode within the mica (ballistic microphonics), it is detected when, with no signal input to the stage, a rise in output current results from tapping the tube. This type of vibration effect is rarely caused by original hole-insert diameters but can be due to the enlargement of the opening by excessive heat treatment of the cathode or a life-cycle requirement of the cathode such that the opening in the mica is enlarged by "puffing" of the mica. Vibration itself can also release small amounts of gas from the puffed mica. This gas can contribute a variable-positive-ion gas current in the grid circuit and, so, create a noise effect.

CONCLUSION

It is apparent that the emission obtainable from an alkaline-earth-oxide film on a nickel-alloy base metal is of the same order of magnitude as that obtained from a cathode system made from a relatively thick coating of oxide. The work functions are essentially the same but the efficiency in terms of milliamperes output per watt input is higher for the thick oxide coating by reason of the porosity factor. The thick oxide matrix coating acts as a reservoir for a replenishment rate of supply of active barium centers by reason of the chemical reducing activity of the base metal to offset the evaporation of barium/barium oxide. The several instances of pulse-emission decay and diminishing performance capability with life are only manifestations of the dynamic equilibria existent in the several physical-chemical processes that exist in the cathode system. The effects of vapor phase interaction and deposition of material onto electrodes in the tube system create unwanted side-reactions. Nonetheless, the oxide-coated cathode is an efficient, consistent performer within those limitations and can be made more stable in operation when due considerations are given to chemical purity in parts processing and sealex operation, temperature limitations, and duty cycle in the end-use application.

Theories of the mechanism of electron emission from a hot metal surface offer an explanation in terms of a kinetic velocity distribution of electrons from the "free mobile" electrons from the valence band into the conduction band. However, a mechanism for transferring electrons through the coating of an oxide cathode is not as yet fully described. The complex nature of the composite surface that is deficient in barium/barium oxide offers a reduced surface-energy-potential barrier and, as such, lowers the work function of the system. In addition to the distribution of barium donor equivalents through the bulk oxide causing the coating to have a high conductivity as a semiconductor material, the porosity of the oxide matrix introduces a factor of electron pore gas that increases the conductivity to a still greater degree, thereby, yielding the high emission performance. The complete mechanism will, ultimately, include the several rate processes, both chemical and physical,[64,65,66,67] that are involved: rate of diffusion of reducing elements in the nickel base metal; rate of reaction with the oxide coating; rate of thermal diffusion of barium; rate of electrolytic transport; and the equilibrium phases of barium (solid) to barium (gas) to barium (ion) forms.

REFERENCES

1. Eisenstein, A. S., Oxide Coated Cathodes, Advances in Electronics, Vol. I, 1948
2. Nichols, M. H., "Thermionic Constants of Tungsten as a Function of Crystallographic Direction," Phys. Rev., Vol. 57, p. 297, 1940
3. Pearson, R. F., "Production of Single Crystal Nickel with Faces Parallel to the Desired Plane," Brit. Jour. Appl. Phys., Vol. 4, p. 342, 1953
4. Loosjes, R., and H. J. Vink, "Properties of Pore Conductors," Philips Research Rpt., Vol. 4, p. 449, 1949
5. Loosjes, R., and C. G. Jansen, "Abnormal Distribution of Velocities of Electrons in Oxide Cathodes During Pulse Emission," Le Vide, Vol. 7, p. 1131, 1952
6. Jansen, C. G., R. Loosjes, and K. Compaan, "Velocity Distribution of Electrons of Pulsed Thermionic Emitters," Philips Research Rpt., Vol. 9, p. 241, 1954
7. Nergaard, L. S., "Studies of the Oxide Cathode," RCA Review, Vol. 13, p. 464, 1952
8. Nergaard, L. S., "Physics of the Cathode," RCA Review, Vol. 18, p. 486, 1957
9. Nergaard, L. S., "Thermionic Emitters," RCA Review, Vol. 20, p. 191, 1959
10. Code R1
11. Plumlee, R. H., "Electrolytic Transport Phenomena in the Oxide Cathode," RCA Review, Vol. 17, p. 190, 1956
12. Plumlee, R. H., and L. P. Smith, "Study of the Sublimation Characteristics of Oxide Cathode Material by Mass Spectrography," Jour. Appl. Phys., Vol. 21, p. 811, 1950
13. Aldritch, L. T., "Study of the Evaporation Products of Barium Oxide from Various Base Metals by Mass Spectrography," Jour. Appl. Phys., Vol. 22, p. 1168, 1951
14. Pelchowitch, I., "Mass Spectrometric Analysis of Evaporation Products of Alkaline Earth Oxides," Philips Research Rpt., Vol. 9, p. 42, 1954
15. Metson, G. H., Electrical Life of an Oxide Cathode Tube Advances in Electronics and Electron Physics, Vol. 8, 1956
16. Wright, D. A., "Survey of Current Knowledge of Thermionic Emitters," Proc. IEE, Vol. 100, III, p. 125, 1953
17. Code R2
18. Code R3
19. Hensley, E. B., "Conduction Mechanism in Oxide Cathodes," Jour. Appl. Phys., Vol. 23, p. 1122, 1952
20. Young, J. R., "Electrical Conductivity and Thermo-electric Power of Oxide Cathodes," Jour. Appl. Phys., Vol. 23, p. 1129, 1952

21. Forman, R., "Measurement and Theoretical Study of Electrical Conductivity and the Hall Effect in Oxide Cathodes," Phys. Rev., Vol. 96, p. 1479, 1954
22. Biguenet, C., and C. Mano, "Emission Carbonates for Vacuum Tube Cathodes," Le Vide, Vol. 1, p. 137, 1946
23. Violet, F., and J. Riethmueller, "A Study of the Oxide Coated Cathode," Annales de Radio-electricity, Vol. 4, p. 184, 1949
24. Code R4
25. Sproull, R. L., "An Investigation of Short Time Thermionic Emission from Oxide Cathodes, "Phys. Rev., Vol. 67, p. 166, 1945
26. Coomes, E. A., "The Pulsed Properties of Oxide Cathodes," Jour. Appl. Phys., Vol. 17, p. 647, 1946
27. Matheson, R. M., and L. S. Nergaard, "The Decay and Recovery of the Pulsed Emission of Oxide Coated Cathodes," Jour. Appl. Phys., Vol. 23, p. 869, 1952
28. Hannay, N. B., D. MacNair, and A. H. White, "Pulsed Emission Levels of Cathode Systems," Jour. Appl. Phys., Vol. 24, p. 1335, 1953
29. Lander, J. J., "The Phase System: Barium Carbonate - Barium Oxide, "Jour. Amer. Chem. Soc., Vol. 73, p. 5893, 1951
30. Lander, J. J., "Experimental Heat Contents for BaO, SrO, and CaO," Jour. Amer. Chem. Soc., Vol. 73, p. 5794, 1951
31. Huber, H., and S. Wagener, "The Crystallographic Structure of Alkaline Earth Oxide Emission Mixtures," Zeit. fur Tech. Physik, Vol. 23, p. 1, 1942
32. Terada, J., "Crystal Structure of the Barium, Strontium, Calcium Triple Carbonate," Jour. Phys. Soc. Jap., Vol. 8, p. 158, 1953
33. Terada, J., "X-ray Study of the Ba-Sr-Ca Triple Oxide," Jour. Phys. Soc. Jap., Vol. 10, p. 555, 1955
34. Ostapchenko, E. P., "X-ray Investigation of the System of Double and Triple Carbonates and Oxides of the Alkaline Earth Metals, "Izvest. Akad. Nauk. S.S.S.R., Ser. z 20, 755, 1105, 1956
35. Eisenstein, A. S., "Study of the Oxide Cathode by X-ray Diffraction," Jour. Appl. Phys., Vol. 17, pp. 434, 654, 1946
36. Faivre, R., and G. Chandron, "X-ray Study of the Structure and Thermodynamic Decomposition of the Mixed Carbonates of Barium, Strontium, and Calcium," Compt. Rend., Vol. 226, p. 249, 1948
37. Shimaza, J., "Emission and Crystal Size of the Oxide Coated Cathode," Jour. Phys. Soc. Jap., Vol. 6, p. 479, 1951
38. Yamaka, E., "X-ray Diffraction Study of Oxide Cathodes," Jour. Appl. Phys., Vol. 22, p. 1087, 1951
39. Yamaka, E., "X-ray Diffraction Studies of Oxide Cathodes," Jour. Appl. Phys., Vol. 23, pp. 498, 937, 1952
40. Eisenstein, A. S., and P. N. Russel, "Thermionic Emission and Electron Diffraction from Thin Films of Barium Oxide, "Jour. Appl. Phys., Vol. 25, p. 954, 1954
41. Grey, L. E., "Pulsed Emission from the System (BaSrCa)O," Nature, Vol. 167, p. 522, 1951
42. White, A. H., "Application of Thermodynamics to Chemical Problems Involving the Oxide Cathode," Jour. Appl. Phys., Vol. 20, p. 856, 1949
43. Rittner, E. S., "Theoretical Study of the Chemistry of Oxide Cathodes," Philips Research Rpt., Vol. 8, p. 184, 1953
44. Hughes, R. C., P. P. Coppola, and H. T. Evans, "Chemical Reactions in Barium Oxide and Tungsten Emitters," Jour. Appl. Phys., Vol. 23, p. 635, 1952
45. Cayless, M. A., and B. N. Watts, "Reactions Occurring During the Decomposition of Alkaline Earth Carbonates," Brit. Jour. Appl. Phys., Vol. 7, No. 10, p. 351, 1956
46. Hoffman, R. E., F. W. Pikus, and R. A. Ward, "Self Diffusion in Solid Nickel," J. Metals, Vol. 8, No. 5, p. 483, 1956
47. Swalin, R. A., and A. Martin, "Solute Diffusion In Nickel Base Substitutional Solid Solutions," J. Metals, Vol. 8, No. 5, p. 567, 1956
48. Peterson, R. W., D. E. Anderson, and W. G. Shepherd, "Influence of the Cathode Base Metal on the Chemical Activation of Oxide Cathodes," Jour. Appl. Phys., Vol. 28, p. 22, 1957
49. Wolk, B., and J. G. Buck, "The Relationship Between Thermal Decomposition in Vacuum and Macrostructure of the Alkaline Earth Carbonates," The Electrochemical Society, p. 190, 1957
50. Claasen, A., and C. F. Veeneman, "Rate of Evaporation of the Alkaline Earth Oxides," Z. Phyzik, Vol. 80, p. 342, 1933
51. Herman, G., and S. Wagener, The Oxide Coated Cathode, Barth, Leipsig, 1943
52. Honig, R. E., "Vapor Pressure Curves for Some Elements," RCA Review, Vol. 18, p. 195, 1957
53. Higginson, G. S., "The Effect of Sulfur and Oxygen on the Electrical Properties of the Oxide Cathode," Jour. Appl. Phys., Vol. 8, No. 4, p. 148, 1957
54. Grattidge, W., and H. John, "Electronic Properties of Barium Sulfide," Jour. Appl. Phys., Vol. 23, No. 10, p. 1145, 1952
55. Code R5
56. Wright, D. A., "The Effect of Coating Core Interface on Conductivity and Emission of Oxide Cathodes," Proc. Phys. Soc. Lond., Vol. A190, p. 394, 1947
57. Eisenstein, A. S., "Properties of the Barium Orthosilicate Interface, "Jour. Appl. Phys., Vol. 20, p. 776, 1949
58. Code R6
59. Haas, G. A., and J. T. Jensen, "Preconversion of Oxide Cathodes, "Naval Research Lab. Bul. No. 5275, PB151442, Office of Technical Services, Dept. of Commerce, Washington 25, D. C., 1959
60. Schrader, E. R., "Survey of Methods Used to Determine Contact Potential," RCA Review, Vol. 18, p. 243, 1957
61. Thompson, B. J., D. O. North, and W. A. Harris, "Fluctuations in Space Charge Limited Currents at Moderately High Frequencies, "RCA Review, Vol. V, No. 3, p. 269, 1940; Vol. V, No. 2, p. 244, 1940; and Vol. V, No. 1, p. 115, 1941
62. Tomlinson, T. B., "Partition Components of Flicker Noise," Jour. Brit. Inst. Radio Eng., Vol. 14, No. 11, p. 515, 1954
63. Schwantes, R., H. J. Hannan, and A. Van der

Ziel, "Flicker Noise in Secondary Emission Tubes and Multiplier Photo Tubes," Jour. Appl. Phys., Vol. 27, p. 573, 1956
64. Peterson, R. W., D. E. Anderson, and W. G. Shepherd, "Influence of the Cathode Base on the Chemical Activation of Oxide Cathodes," Jour. Appl. Phys., Vol. 28, pp. 22-31, 1956
65. Noelcke, C. L., "Deterioration Mechanisms in Electron Tubes," ARINC Research Monograph No. 6, Publication No. 101-4-127, 1958
66. Code R7
67. Code R8

Nickel Base Metal for the Oxide Cathode

C. H. Meltzer and E. G. Widell

Harrison

The oxide-coated cathode is more than an electrode for conducting current in an electron tube. It has the additional and most important function of providing the efficient transfer of electrons for emission that makes conduction possible in the tube. The emission is the result of a dynamic equilibrium system based on the interaction of reducing agents in the base nickel with the coating of alkaline-earth oxides. In such a system, the activated coating, consisting of barium oxide and other oxides, is considered to be a semiconductor containing active donor sites associated with the presence of free barium. The existence of such an activated coating on the substrate nickel alloy metal forms a low work function system that permits efficient transfer of electrons from the cathode at a relatively low temperature of 1025 K. Barium is most readily formed in the coating by the reduction of the barium oxide component of the coating by reducing agents such as carbon, magnesium, and silicon supplied by the supporting nickel alloy metal. The kind and amount of the specific reducing elements added to the cathode base metal depend upon the end result required of the specific tube design as determined by the tube design engineer.

NICKEL

Nickel is the metal accepted by the tube industry as most suitable for the cathode base material. Although nickel has much in common with iron and cobalt, and with them, forms a family (group VIII) in the periodic table of elements,[1] nickel has the best combination of characteristics for base-metal purposes (see Table I). Nickel has good elastic and mechanical properties; it does not oxidize easily and forms relatively unstable oxides, and therefore, is readily deoxidized by heating in hydrogen (hydrogen firing); it is ductile and, as a result, is capable of being formed into the shapes required for cathode applications. In addition, nickel can take a high polish with a resultant low surface emissivity, i. e., it has low radiant heat losses; its heat conductivity is a good compromise between enough to insure uniform cathode sleeve temperature and not so much as to cause excessively high end losses; its electrical conductivity is adequate, about one-fourth that of copper; and most important, its high melting point, low vapor pressure, and relative freedom from oxidation make it possible to process and operate tubes at the cathode temperatures needed to obtain suitable emission capability.

REQUIREMENTS OF NICKEL ALLOYS FOR CATHODE BASE MATERIAL

Early in the manufacture of oxide-coated cathodes it was discovered that just any nickel was not suitable for use as a cathode base material. Early oxide-coated cathodes consisted of a nickel filament which was coated with alkaline-earth compounds or which had the compounds incorporated in the filament material. Investigations disclosed that the presence of barium in the oxide coating affected the emission and that barium could be made available by the reaction between the barium oxide in the coating and certain reducing agents in the base metal. The problem, therefore, was to decide what agents gave the best emission performance, and also, to determine what additives would achieve the desirable result of increasing the hot resistivity and the hot strength of the filaments. Many alloys were tried but the achievement of an alloy having all the desired improvements was limited by unwanted side reactions introduced by some of the additives. These side reactions will be described later.

The filament type of cathode was largely displaced in receiving types of tubes by the indirectly heated unipotential type of cathode which offered many advantages for ac heater-power supply. In the unipotential type of cathode, the fact that the heater function and the coating-support function are separate has changed the requirements for the physical properties of the nickel alloy. Its hot strength and hot resistivity are no longer the major considerations and, therefore, full consideration can be given to those properties best suited to a coating support material and to the reducing element content needed for suitable activation of the oxide coating.

As a practical matter, the selection of suitable compositions for the cathode base metal has been influenced by the availability of the alloys in commercial use. The varying properties of these alloys produced by the air-melting processes employed by the nickel industry made it mandatory to select individual ingots which gave compositions most favorable to achieving the desired tube performance and uniformity. The success of this method over the past twenty years has depended on the availability of these better alloys as well as economic factors of cost and demand. The need of the tube industry for nickel represents but a very small portion of the tonnage made for the steel and allied construction industries. In recent years, the situation has improved in that nickel alloys produced by a vacuum-melt process have been available. As early as 1949, RCA demonstrated that nickel wire for filament-type cathodes produced by vacuum-melting techniques had superior qual-

ities for filaments to wire produced by air-melting processes. In 1958, RCA began large-scale production of vacuum-melt nickel alloys for indirectly heated cathode sleeves. Large scale production (melts of 500 pounds) was called for in order to insure the correctness of the desired compositions by reducing the variations between melts and within melts which often occurred with alloys purchased from commercial vendors.

Table I

Physical Properties of Nickel[1]

Atomic number		28
Atomic weight	gm	58.69
Atomic volume	cc	6.59
Density	gm/cc	8.9
Melting point	C.	1452
Boiling point	C.	2730
Electrical resistance	ohms x 10^{-6}	6.9
Curie point (loss of magnetism)	C.	360
Tensile strength	lb/in^2 cast	53000
	drawn	150000
Metal to ion plus 2 electrons, oxidation potential		0.25
Ionization potential of gas atom	1st electron	7.61
	2nd electron	18.22
Specific heat	cal/gm at 500 C.	0.127
Thermal conductivity	cal/cm sec deg C. at 500 C.	0.148
Thermal emissivity		0.355
Work function in electron-volts		4.98

METALLURGY OF NICKEL AND ITS ALLOYS

Nickel in ores almost always occurs when associated with cobalt, iron, and copper as complex minerals of sulfur and arsenic, as well as in the form of silicates. The most important nickel-iron-copper-bearing sulfide ore is found in Ontario, Canada; considerable lower-grade nickel-magnesium silicate ores are found in New Caledonia and the West Indies. The International Nickel Company of Canada, Ltd. is the chief producer of nickel. The metallurgy of nickel is complex; the method that is used depends upon the nature of the ore.[2] The final product of the processing is nickel oxide containing impurities of sodium, copper, zinc, lead, sulfur, and arsenic. The nickel oxide is reduced to nickel with soft coal, melted in a reverbatory furnace, and poured into molds. The nickel ingots produced by this process are 96 per cent pure. These ingots are then used as anodes in an electrolytic cell having nickel sulfate as the electrolyte. The pure nickel deposited at the cathode is the electrolytic nickel used for the manufacture of nickel alloys for cathodes. This electrolytic nickel contains 99.95 per cent of pure nickel and cobalt. The range of the cobalt content is from 0.3 to 0.5 per cent. Faint spectrographic trace amounts of copper, iron, silicon, carbon, and sulfur are present. Another form of pure nickel is made for use in the powder metallurgy technique. In this instance, the nickel oxide is reduced with a mixture of hydrogen and carbon monoxide (water gas) at 300 C. The nickel is purified as follows: first, the carbonyl gas compound is made by passing a stream of carbon monoxide over the nickel powder at 80 C to form nickel carbonyl gas, $Ni(CO)_4$; then, the gas is decomposed at 200 C in tall stainless steel towers to yield the pure nickel powder. The particle size of the nickel powders is dependent upon the rate of flow of the nickel carbonyl gas through the heated towers and is smallest when the rate of gas flow is slowest. This carbonyl nickel contains 99.95 per cent of pure nickel and cobalt; the cobalt content is about 0.3 per cent. The remainder of the powder is about 0.06 per cent carbon.

The next step is to melt the nickel in either its electrolytic form or carbonyl nickel powder form into ingots suitable for further treatment of the metal to form the final cathode nickel alloy. Two melting procedures are in commercial use; the air-melt process and the vacuum-melt process. These methods will be discussed in turn.

Air-Melt Process

In the air-melt process, electrolytic nickel is melted in acid-type, natural-gas fired reverbatory furnaces with capacities of 6 to 20 tons, or in basic-type induction furnaces with capacities of 1 to 3 tons. For the acid process, calcium silicate lines the side walls and sintered quartz sand forms the bottom lining; for the basic process, the linings are made from calcium carbonate (limestone) and magnesium carbonate (magnesite). The charge is clean mill scrap from previous runs and essentially sulfur-free, electrolytic nickel. Charcoal, with a maximum of 1 per cent sulfur content, is used to deoxidize the metal. The furnace gases are essentially sulfur-free and are made slightly reducing by the addition of about 2 per cent carbon monoxide. Excess carbon in the molten metal is removed by reacting it with nickel oxide added to the charge. Finishing amounts of magnesium and manganese are added to ensure complete deoxidation. The molten metal is then poured into copper-base molds for cooling. These ingots of cathode nickel alloys are forged into flats or bars which can be subsequently processed into strip, rod, or wire by hot-rolling and cold-working operations.

The pure nickel is made malleable by adding selected additives so that the ingot can be hot-rolled and cold-worked to the desired strip gauge. The process involves four essential requirements: deoxidation to remove harmful oxides; degassification to remove carbon dioxide, carbon monoxide, nitrogen, and hydrogen; fixing of harmful elements by compound formation; and the avoidance of any injurious effect from the selected additives. No single addition element is capable of satisfying all four requirements. Aluminum will deoxidize, but will not fix harmful elements. Boron acts as a fluxing agent, improves the fluidity of the molten metal, and combines with the last traces of gas that is not fixed by other deoxidizers; an excess of it will cause brittleness. Calcium will deoxidize, as well as fix sulfur, but an excess of it will cause brittleness. Magnesium is sufficient for purposes of deoxidation and can fix sulfur but it can not provide adequate degassification. It will not fix selenium or tellurium completely. Magnesium alone is inadequate for the more severe hot-working operations such as rolling sharp-cornered sections or hot-piercing of seamless tubing. It produces material that can be hot-worked well at high tem-

peratures but which "edge-cracks" at low temperatures. Manganese is used to deoxidize the molten nickel as well as to fix sulfur. It improves the rolling and forging qualities of the nickel and contributes some hardening properties; an excess of it results in a loss of ductility and machinability. Silicon is used to deoxidize the nickel but it will not fix harmful elements. Titanium will convert some harmful solid and gaseous nitrogen elements but will not provide adequate deoxidation. Phosphorous will convert an excess of calcium to a harmless constituent but will not remove gases; an excess of phosphorus renders the nickel brittle at low temperatures.

Fundamental studies of the processing involved in the fabrication of nickel and nickel alloys have furnished adequate proof that melting of nickel must be accomplished under conditions that remove or prevent the introduction of oxides or sulfides into the cast metal. Deoxidation as such is only part of the treatment which would give the metal adequate hot and cold malleability for all processing and fabricating operations. It is concluded that a small amount (0.01 per cent) of sulfur renders nickel nonmalleable because of the formation of nickel sulfide (Ni_3S_2) which surrounds each metallic grain of nickel.[3] This sulfide lowers the intercrystalline cohesion of the mass of nickel grains even at low temperatures. Nickel sulfide melts at 630 C; consequently, at temperatures above this critical point, there is practically no cohesion between the grains of nickel.

As previously stated, the treatment for producing optimum malleability of nickel alloys requires the use of two or more addition elements. Ingots of nickel alloys for cathode purposes have been deoxidized mainly by the addition of controlled amounts of carbon, magnesium, and manganese because the other deoxidizing elements in excess of critical amounts would impair the performance of the alloy as a cathode base metal in electron tubes. The amount of each element that is used to produce such optimum malleability must be varied to suit the chemical composition of the nickel material that is being treated and its condition in the air-melting furnace at the time of treatment. If excess sulfur is present in the nickel, the material can not be worked at temperatures below 925 C without fracture along the edges. Calcium, which may appear in the molten metal because of the contact between the molten metal and the refractory lining of the furnace, affects the upper range of hot ductility. The excess calcium must be counteracted by a material that forms a compound with it, e. g., boron or phosphorus. Elements that can cause difficulties in the hot-working of these nickel alloys are: calcium in excess of 0.001 per cent; lead in excess of 0.02 per cent; sulfur in excess of 0.01 per cent; selenium or tellurium in excess of 0.001 per cent; boron in excess of 0.001 per cent; and phosphorus in excess of 0.002 per cent.

In practice, the sulfur content of nickel alloys should be substantially below 0.008 per cent, because sulfur in excess of only 0.01 per cent will render the nickel nonmalleable. The addition of manganese to react with the sulfur slightly improves the malleability of the metal. The resulting manganous sulfide MnS forms a eutectic with nickel that melts at 1325 C. It is distributed in the form of a network of small globules along the grain boundaries of the nickel lattice. Manganese is used, however, despite its slight ability to fix sulfur, because it does perform a necessary amount of deoxidation and contributes to improved forging qualities of the nickel at high temperatures, i. e., it helps to produce ingots with clean surfaces free of "pulls" and "cracks" during the rolling and drawing operations. Merica and Waltenberg[3] have indicated that the addition of magnesium to the molten nickel reacts with the sulfur to form magnesium sulfide MgS, which is insoluble in the nickel and has a melting point higher than that of nickel. The magnesium sulfide forms small crystalline particles which are entrapped during the solidification of the nickel and become dispersed throughout the grains but not in the grain boundaries (see Fig. 1). At this stage, the metal is malleable. However, nickel containing sulfur and made malleable by the addition of magnesium becomes nonmalleable when the alloy is melted in an oxidizing atmosphere. The magnesium sulfide decomposes to form objectionable films of nickel sulfide around the nickel grains. Nickel oxide does not enter into the grain boundaries as nickel sulfide does (see Fig. 2). It is considered that the solubility of nickel oxide in solid solution in nickel is slight. The eutectic form, Ni-NiO, exists up to 1.1 per cent nickel oxide; in excess of this percentage composition, the nickel oxide separates out uncombined. It is concluded that nickel oxide has little or no effect upon the malleability of the nickel.

The composition specifications for some commercially available air-melted nickel alloys suitable for cathode base metal are shown in Table II by RCA alloy numbers N18, N34, N81, N109 and N130. These figures are maxima or limiting ranges of composition in weight per cent as applied to final strip or tubing; the cathode sleeve formed from these alloys will have a much lower carbon content after it has been hydrogen fired during the cleaning process. The analysis does not differentiate between the effective free element content and the total combined content including oxide and intermetallic compound formation. The figure of merit is given in terms of total content. The unknown distribution with respect to the free element (Mg), its oxide (MgO), and compound form (MgSi or Mg_2C_3), for example, does not permit accurate estimates of the reducing element activity in terms of cathode effectiveness. Hence, the selection of cathode base metal has been based on pilot-test tube performance in lieu of a better means for evaluating the activity of the nickel alloy. Air-melt cathode nickel possesses a further complicating factor caused by the lack of homogeneity from melt to melt as well as within a given melt. Nonhomogeneity within the melt results from the large tonnage melts used by the commercial producers. The large molten mass of nickel cools down very slowly. As a result, segregation of additives occurs because pure nickel crystallizes out first, and the additives are therefore forced to diffuse out in front of the solid nickel phase into the remaining liquid zone. Subsequent solidification causes the additives to be present in random pockets of concentration. The nonhomogeneity between melts results from the acceptance of in-grade material selected on

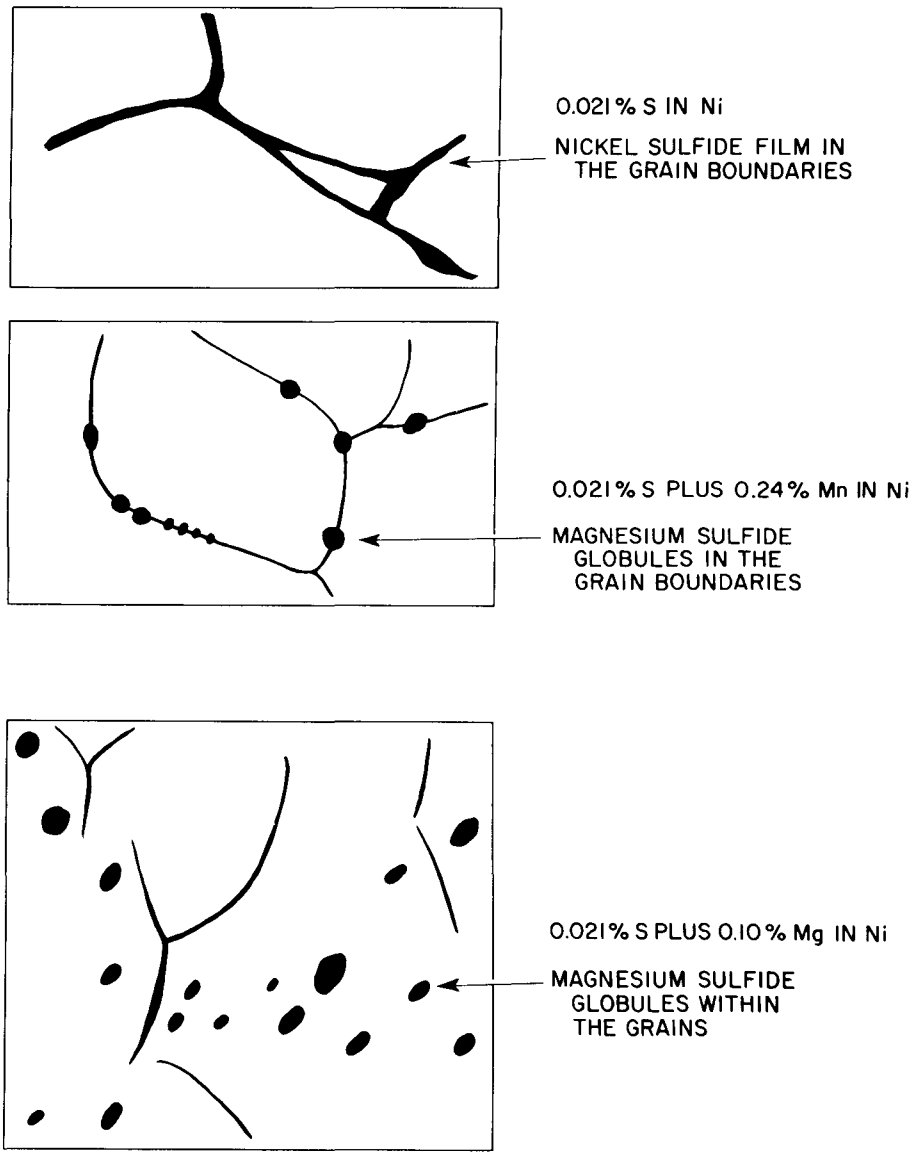

0.021% S IN Ni
NICKEL SULFIDE FILM IN THE GRAIN BOUNDARIES
0.021% S PLUS 0.24% Mn IN Ni
MAGNESIUM SULFIDE GLOBULES IN THE GRAIN BOUNDARIES
0.021% S PLUS 0.10% Mg IN Ni
MAGNESIUM SULFIDE GLOBULES WITHIN THE GRAINS

Figure 1. Photomicrographs of Sulfur in Nickel (500X)

the basis of broad composition specifications which the vendor can meet on a tonnage scale.

Vacuum-Melt Process

The requirements for improving the stability, reliability, and life of tubes in electronic applications created the demand for better cathode base-metal alloys. Controlled compositions designed specifically for good emission performance with long life and minimum side reactions such as high interface impedance, excessive leakage paths along insulating surfaces, grid emission, and reverse emission from the anode were desired. Such special alloys for cathode purposes were not available until RCA began its development of vacuum-melting techniques which would insure that the additives, and particularly those in trace amounts, in the compositions could be maintained within narrow limits. The use of vacuum-melting techniques for producing cathode-grade nickel alloys enables a much higher degree of homogeneity and purity to be obtained than is possible with conventional air-melting techniques. Because contamination by the furnace and its natural-gas atmosphere is avoided, it is unnecessary to make any specific additions of reducing elements for purposes of deoxidation and fixing of harmful ingredients. The composition of the nickel alloy in terms of reducing element activity is more closely related to the requirements of cathode base metals than to the demands of air-melt metallurgy. However, because hot-rolling of the ingots is carried out in the oxidizing atmosphere of air, and cold-rolling techniques require specific degrees of malleability, certain additives, such as manganese, are incorporated as for the air-melt technique. The need of hot-strength nickel alloys for some cathode applications, where bowing of the sleeve must be minimized at high temperature operations, requires the presence of other additives that can exert troublesome side reactions when they are present in excess of critical amounts.

The reproducibility of vacuum-melt cathode nickel-alloy compositions is based on the use of electrolytic

nickel of 99.95 per cent purity nickel-cobalt (the range of cobalt content is about 0.3 per cent); the addition of precise amounts of chemically pure reducing element additives; and a process which prevents the introduction of foreign impurities. The problem of homogeneity within the melt is resolved by limiting the ingot size to two 250-pound portions that are poured from a 500-pound melt. The 500-pound batch is small enough to be poured quickly into the molds; each ingot is small enough to solidify rapidly and practically eliminate the segregation of additives since there is a minimum of differential cooling within the molten mass of the ingot.

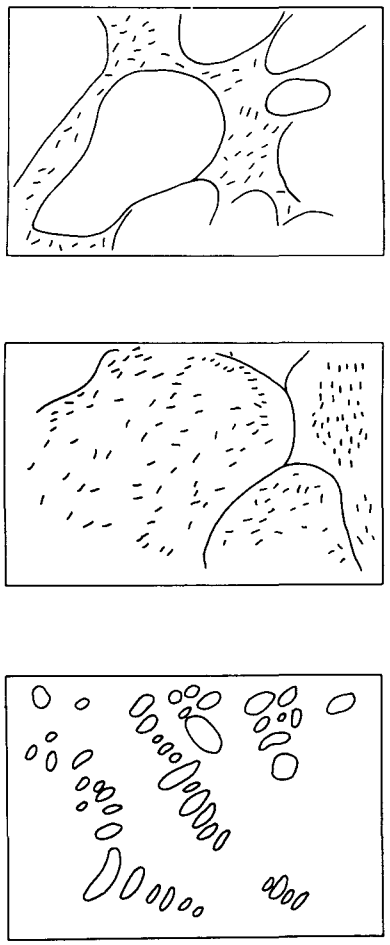


Figure 2. Photomicrographs of Nickel Oxide in Nickel (100X)

The charge of electrolytic nickel is melted in a magnesia crucible in a vacuum of less than 25 microns of mercury. The crucible is lined with a cement facing of magnesia-alumina that is chipped out between changes of composition in order to prevent contamination of one charge by the other. When the nickel has melted, hydrogen gas is admitted into the system and maintained at a pressure of 20 millimeters of mercury in order to deoxidize the molten nickel. After a while, the hydrogen and the resultant water vapor are pumped away. The molten nickel is then kept under an inert atmosphere of argon gas maintained at some 250 millimeters of mercury. At this time, the selected reducing elements are added to the molten nickel in a definite order, at precise intervals and at temperatures and pressures to minimize the effects of intermetallic compound formation, and to hold losses of the volatile ingredients, like magnesium, to some constant value. The total amount of reducing element agents that are added is adjusted to compensate for such losses so that the desired amount is obtained in the final alloy. Cobalt, tungsten, silicon, manganese, and carbon do not require compensation amounts; magnesium is added as an alloy of nickel to minimize its losses. The amount of magnesium alloy added is adjusted so that the amount of magnesium remaining in the melt is a constant under the particular conditions of time, temperature, and pressure. Carbon is added in excess of the desired amount, not because of losses incurred during the melting operation, but to allow for losses which will be incurred during the hot-rolling of the ingot and the final hydrogen firing of the cathode. The final amount of carbon that is present in the finished cathode sleeve is held to close limits. The presence of carbon indicates that the nickel alloy has not been oxidized too greatly during the rolling operations, and thus, indicates that the reducing element content has not been essentially changed by oxide compound formation. However, when too much carbon is present in the final cathode after hydrogen firing, it can cause gassy tubes during sealex manufacture of the tube by the reaction of carbon with barium oxide to form carbon monoxide. The final carbon content in the cathode sleeve is held to a minimum amount within close limits to insure the optimum result. The specifications for vacuum-melt cathode alloys are illustrated in Table II by RCA alloy numbers N132 and N136.

EFFECTS OF REDUCING AGENTS ON HOT STRENGTH AND HOT RESISTIVITY

While increased hot strength of nickel alloys has been one of the primary problems for directly heated cathodes (filamentary type), current design of tube types to withstand shock and vibration has made increased nickel-alloy hot strength important for indirectly heated cathodes. The tube assembly is trussed in a rigid framework so that the heated cathode sleeve is kept from expanding through the mica supports. Under these conditions, bowing of the cathode sleeve can occur unless the nickel alloy has sufficient hot strength to withstand the pressures created by the restricted expansion movement. Considerations of some of the data accumulated on experimental melts indicates that intermetallic compounds may play an important role in the hot strength of nickel alloys. The effects of reducing-agent additives in the nickel alloy for increasing the hot strength of the metal may be summarized as follows:[4]

1. Magnesium alone: little effect
2. Silicon alone: little effect
3. Carbon alone: little effect
4. Aluminum alone: little effect
5. Aluminum plus silicon but no carbon: little effect
6. Aluminum plus carbon: slight effect
7. Magnesium plus carbon: great effect
8. Silicon plus carbon: appreciable effect
9. Zirconium plus carbon: great effect
10. Titanium plus carbon: very great effect
11. Tungsten plus carbon: very great effect

Table II

Domestic Cathode Nickel Alloys

Limiting compositions* on tubing or finished strip

(maximum except where range is given)

Manufacturer's Designation**				Indirectly Heated Cathodes											
Inco	DH	ST	RCA	C	Cu	Fe	Mg	Mn	Si	S	Ti	Al	W	Co	Ni-Co
-	-	-	N2S	.04	.05	.05	.02-.08	.05-.15	.10-.16	-	0	-	-	38-42	-
330	-	G10	N7T	.10	.15	.20	pres.	.30	.10	.008	.005	-	-	-	99.0
220	-	G11	N15-N18-N18T	.04	.20	.20	.01-.10	.20	.01-.05	.008	.03	-	-	-	99.2
225	-	G3	N19-N34	.04	.20	.20	.01-.08	.20	.15-.25	.008	.03	-	-	-	99.0
-	N42	-	N42S	.08	.05	.05	.05-.10	.05-.10	.15-.20	.005	.005	0	-	.4-.7	bal.
-	399	G6	N80-N116	.04	.10	.10	.005	.10	.15-.25	.008	.01	-	-	-	99.4
-	499	G21	N81-N81T	.02	.10	.10	.005	.02	.010	.005	.005	-	-	-	99.8
			N107	.04	.05	.05	.02-.08	.05-.15	.02-.05	-	0	-	-	38-42	bal.
Selected 330	-	-	N109S & T	.08	.15	.10	.03-.10	.20	.01-.05	.008	.005	-	-	1.0	99.0
-	699	G5	N120T	.04	.10	.10	.10	.10	.05-.12	.008	.02	-	-	-	99.5
-	599	G4	-	.08	.05	.10	.01	.10	.15-.20	.005	-	-	-	-	bal.
-	799	G2	-	.08	-	-	.08	-	.20	-	-	-	-	-	bal.
-	899	-	-	.05-.10	-	.10	.05-.10	-	.02-.07	-	-	-	-	-	bal.
-	999	P50	N131T	.05	.04	.05	.01	.02	.02	.005	.01	.002	-	-	99.5
-	-	A30	-	.03-.10	.05	.10	.02-.07	.05	.02	.005	.01	.07-.13	-	.08	99.2
-	-	A31	N130	.03-.10	.10	.10	.02-.07	.05	.01-.06	.005	.02	.002	3.75-4.25	1.0	94.5
-	-	A32	-	.03-.10	.05	.10	.02-.07	.05	.03	.005	.01	.07-.13	2.0 -2.50	-	96.2
-	-	X3010†	-	.05	.10	.10	.01-.06	.05	.02	.005	.02	-	3.75-.425	-	bal.
-	-	X3012†	-	.03-.08	.05	.10	.01	.10	.01	.005	.01	-	1.75-2.25	-	bal.
-	-	-	N132 (vacuum melted 330)	.01-.03	.10	.10	.02-.04	.08	.01-.03	.008	.005	-	-	.3-.5	bal.
-	-	-	N133 (vacuum melted A31)	.03-.10	.10	.10	.02-.07	.05	.01-.06	.005	.02	.002	3.75-4.25	1.0	94.5
-	-	-	N136 (vac. melt hi-Mg 330)	.01-.03	.10	.10	.04-.07	.08	.01-.03	.008	.005	-	-	.3-.5	bal.
-	-	-	N144 (air melt P51)	.05	.04	.05	.01	.02	.02	.005	.01	.002	3.75-4.25	-	95.5
-	-	-	N702†† (vacuum melt, hi Co)	.01-.03			.07	.01	.01			2.0		40	

Directly Heated (Filamentary) Cathodes

(nominal composition except where range is given)

-	-	-	N1W	.10	.05	.05	.05	-	-	-	0	1.0	2.0	40	-
Selected 225	-	-	N9	-	.20	.20	.02-.07	.20	.12-.20	.008	.01	-	-	1.0	bal.
Cobanic WBD	-	-	N85	.05	low	.90	.04	.03	.18	-		-	-	42-45	-
-	-	-	N91	.04	.05	.05	.01	.05-.15	.075-.125	-	0	-	-	18-22	-
SC	210	-	N94	.34	.03	.07	.20	.05	.18	-	.2	0	-	.2	-
-	-	-	N97	.04	.05	.05	.01	.05-.15	.12-.20	-	0	-	-	38-42	-
-	-	-	N100	.10	.05	.05	.05	-	-	-	0	1.0	2.0	.5	bal.
SC	213	-		-	-	-	-	U-1.0	-	-	-	1.9	1.0	-	96.1

* Zero composition means absent spectrographically or less than 0.005%; dash means "not specified."

** Inco = Int'l. Nickel Company; DH = Driver-Harris; ST = Superior Tube; WBD = W. B. Driver; SC = Sigmund Cohn

† Contains 0.1% Zirconium

†† Contains 5% Molybdenum

The cumulative effect of a combination of elements is generally more effective in raising the strength and hardness of an alloy than an equivalent addition of a single element. The hardenability effects of several elements are multiplicative, while increasing amounts of one element are additive.[5] It is evident that the addition of some of the reducing agents, i. e., silicon, aluminum, and magnesium, alone or in combination has little effect, but that those elements which have a tendency to form stable carbides impart strength to the nickel. Although zirconium and carbon impart great strength to the nickel, the most pronounced increase in hardness occurs with the addition of titanium and carbon. The addition of 0.10 per cent zirconium plus 0.10 per cent carbon increases the tensile strength of pure nickel by 50 per cent, but the same amount of titanium and carbon produces an increase of nearly 100 per cent. The hardening effects of several reducing elements added to nickel are shown in Table III.

Laboratory tests conducted by C. W. Horsting have indicated that the addition of 2 per cent tungsten to

nickel increases the hot strength of the alloy by 20 per cent. Further additions of tungsten increase the hot strength of the nickel alloy at a much slower but uniform rate. The addition of cobalt to the extent of 40 per cent of the alloy seems to have little hardening effect when no aluminum is added. However, when the alloy is made with cobalt plus aluminum by the air-melting process, there is a marked increase in the hot strength. When the same alloy is made by the vacuum-melting process, there is only a slight increase in the hot strength. The increased hot strength of the alloy made of cobalt and aluminum in nickel under air-melt conditions may be due to the formation of microcrystalline grains of aluminum oxide in the slip planes of the nickel metal. Titanium appears to produce the greatest hardening effect of the several additives. It is also an excellent reducing agent, but the side reactions of heavy, dark-brown interface compound formation precludes its use in the amounts necessary to harden nickel.

Table III
Tensile Strength of Some Nickel Alloy Base Metals

Metal	Nominal Composition -- per cent	T.S. in psi After Annealing at 1000 C	
		800 C	20 C
N81	0.02 C	8,200	28,600
N18 (220)	0.03 Mg 0.02 Si 0.09 Co 0.03 Ti	10,300	-
N109 (selected 330)	0.07 Mg 0.03 Si 0.10 Co	13,400	43,600
N2S	0.06 Mg 0.2 Si 40.0 Co	13,700	59,000
N34 (225)	0.20 Si 0.03 Co 0.03 Mg 0.03 Ti	14,200	44,800
N107	0.05 Mg 0.03 Si 40.0 Co	15,000	58,700
N130 (A31)	0.03 Si 0.03 Mg 0.5 Co 4.0 W	17,000	60,900
(Selectron)*	3.0 Si	18,700	-
(16 Metal)*	0.5 Ti 1.5 Si 17.0 Co 0.1 Al	20,700	-
(LX74)	1.0 Al 10.0 W 5.0 Mo 40.0 Co	26,300	96,000
(Konel)*	2.5 Ti 0.5 Si 0.3 Al 17.0 Co	31,300	-
(Tungsten T10)	100.0 W	130,000	-

* Obsolete, coating peeled

The effect of various additives on the electrical resistance of nickel at 20 C is given in Fig. 3 from data obtained from Wise and Schaefer.[6] Additional data on changes of the electrical resistance of nickel by the addition of aluminum and tungsten elements have been supplied by C. W. Horsting.

EFFECTS OF REDUCING AGENTS ON CATHODE ACTIVITY AND REACTIONS

Commercial practice makes use of the reducing element content of the nickel alloy to effectively produce an activated oxide coating by creating barium and associated donor sites in the oxide matrix as a result of the chemical reduction of barium oxide. The reducing agents of carbon, magnesium, silicon, titanium, manganese, and tungsten react with the barium oxide (BaO) component of the emission coating to form barium (Ba) and the respective products of the reaction at the operating temperature of the tube — 980 to 1020 K.

$$BaO + C = Ba + CO \quad \text{(carbon monoxide)}$$
$$BaO + Mg = Ba + MgO \quad \text{(magnesium oxide)}$$
$$4BaO + Si = 2Ba + Ba_2SiO_4 \quad \text{(barium orthosilicate)}$$
$$4BaO + Ti = 2Ba + Ba_2TiO_4 \quad \text{(barium titanate)}$$
$$BaO + Mn = Ba + MnO \quad \text{(manganese oxide)}$$
$$6BaO + W = 3Ba + Ba_3WO_6 \quad \text{(barium tungstate)}$$

The reducing agents added to the nickel play an important role in activating an oxide-coated cathode. Electrolysis plays a supplementary role during aging and life of the tube. If electrolysis were the sole means of activating the cathode system employing very pure nickel base metal — i. e., passive metal — the time required to bring the cathode system to a good level of emission would be about 200 hours. Consequently, tube manufacturing operations must rely on the activity of reducing agents to achieve good initial activation. If the reducing agents should be added to the emission carbonate coating, chemical reactions would occur prior to the decomposition of the carbonates to the oxide form. These reactions take place between the carbonate and the reducing agent at the low temperature of 850 K to yield carbon monoxide plus the metal compound:

$$BaCO_3 + Mg = BaO + MgO + CO$$
$$2BaCO_3 + Si = Ba_2SiO_4 + 2CO$$

The same harmful side reactions can occur when the decomposition of the emission carbonate is prolonged at a slow rate through a relatively low temperature zone below 850 K. In this case, the surface concentrations of reducing agents in the base metal are effectively oxidized. The resultant oxide film, or interface compound of barium orthosilicate when it is present, creates a barrier for further rapid diffusion of reducing elements and, in effect, lowers the activity of the cathode system. As a result, low levels of emission performance are experienced on initial testing and early life.

The activity of the several reducing agents varies in terms of concentration, free energy of compound formation, and duration. The most effective reducing agents

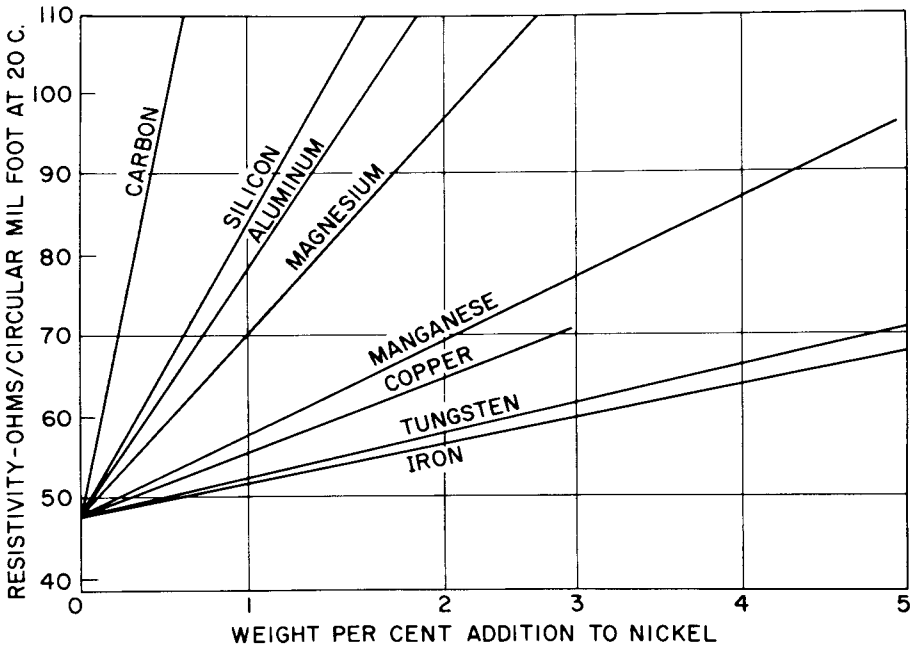


Figure 3. Effects of Various Additives on the Resistivity of Nickel (aluminum and tungsten by C. W. Horsting; all others by E. M. Wise and R. H. Schaefer[6])

may also induce undesirable side reactions. For these reasons, several reducing agents are added to the cathode base metal to obtain a balance which yields good emission together with long life performance and relative freedom from the side effects of sublimation, grid emission, high interface impedance, and changing contact potential.

Barium Vapor Pressure and Diffusion Rates

Equilibrium pressures of barium produced by the reduction of barium oxide have been calculated by White[7] and Rittner[8] in order to evaluate the activity of the several reducing agents. These calculations are based on molecular reactions of the pure reactants in intimate contact with each other. The results can be used only as reference values because the reactants are neither in molecular proportions nor are they as readily available in the oxide cathode system. The concentrations of the reducing agents are minute fractions of molecular amounts; a diffusion zone and an interface zone separate the barium oxide component from the reducing elements. Thus it is necessary to take into account the rate of diffusion of the reducing element to the surface of the cathode metal and then through the interface layers. Table IV lists the activities (in terms of equilibrium barium vapor pressure) of some of the pure-metal reducing elements as calculated by White for 1073 K and by Rittner for 1000 K Table V lists the rate of diffusion of the various reducing elements through nickel as reported by Shepherd, et al.[9] The diffusion rate of carbon has been determined by Widell and Umbreit.[10]

Sublimation Effects

Electrical leakage between electrodes, heater-cathode leakage, and cold-loading effects have been attributed to material subliming from the cathode base metal. The ASTM[11] has standardized a method for studying this phenomenon whereby the material subliming from the uncoated cathode can be collected on a mica target to produce a change in resistance between two electrodes attached to the mica. The use of mica as a target material has not been too successful because variations in the purity of the mica and the extent of its surface variations have resulted in erroneous data. More recently, the use of quartz substrate material faced with platinum electrodes has made possible more reliable data capable of standardization.

Table IV

Equilibrium Pressure of Barium Produced by the Reduction of Barium Oxide[7, 8] (in order of increasing activity of base metal)

Reaction	P_{Ba} in mm Hg	Temperature degrees Kelvin
$BaO(s)+Ni(s) = Ba(g)+NiO(s)$	9×10^{-17}	1073
$3BaO(s)+Mo(s) = 2Ba(g)+Ba_{(s)}MoO_3$	6×10^{-12}	1000
$3BaO(s)+W(s) = 2Ba(g)+Ba_{(s)}WO_3$	2×10^{-9}	1000
$BaO(s)+Mn(s) = Ba(g)+MaO(s)$	1×10^{-8}	1000
$2BaO(s)+Si(s) = 2Ba(g)+SiO_2(s)$	3×10^{-8}	1073
$BaO(s)+C(s) = Ba(g)+CO(g)$	8×10^{-6}	1073
	4×10^{-6}	1000
$2BaO(s)+Ti(s) = 2Ba(g)+TiO_2(s)$	7×10^{-6}	1000
$BaO(s)+Mg(s) = Ba(g)+MgO(s)$	**7**	**1073**

Table V

Diffusion Constants of Reducing Agents in Nickel[9] *

Reducing Agent	Diffusion Constant cm^2/sec	Temperature Range — deg C
Magnesium	$0.44\, e^{-\frac{56,600}{RT}}$	1067-1299
Silicon	$1.5\, e^{-\frac{61,700}{RT}}$	1115-1303
Molybdenum	$3.0\, e^{-\frac{68,900}{RT}}$	1155-1402
Titanium	$0.86\, e^{-\frac{61,400}{RT}}$	1155-1402
Aluminum	$1.87\, e^{-\frac{64,000}{RT}}$	1155-1402
Manganese	$7.50\, e^{-\frac{67,100}{RT}}$	1155-1402
Tungsten	$11.1\, e^{-\frac{76,800}{RT}}$	1155-1402
Carbon	$0.456\, e^{-\frac{39,300}{RT}}$	700-1000

*Carbon by Widell and Umbreit[10]

Most complaints of electrical leakage have come from the use of nickel alloys containing more than 0.08 per cent of magnesium. The reason that magnesium is cited as being the main source of leakage effects is based on an examination of the vapor pressure curves of Fig. 4[12]. Magnesium has a vapor pressure nine orders of magnitude greater than that of nickel at 1150 K. It also has a very high diffusion rate in the cathode nickel alloy. The amount of magnesium that does sublime is dependent on the factors of concentration, rate of diffusion, and the temperature reached during the manufacturing process. The deposits on adjoining electrodes, micas, and glass that result from such sublimation have been analyzed spectrographically as well as by wet chemical methods. The deposit is shown to be mostly magnesium; manganese and nickel have also been found in smaller amounts.[13] The nickel found on mica spacers results from a high rate of evaporation from bare portions of the cathode operating at or above 1150 K (brightness temperature). Ordinarily, the oxide coating prevents sublimation from the coated portions of the nickel cathode unless the operating temperature is in excess of 1200 K. The operating temperature ranges of some cathode systems are shown in Table VI.

REVIEW OF SPECIFIC REDUCING AGENTS

The effects of specific reducing agents in the nickel alloys will be considered for the following eleven additives: carbon in pure nickel - RCA No. N81 (Driver-Harris 499 alloy); carburized nickel alloy - RCA No. N18 (Inco 220 alloy); carbon and magnesium; magnesium; silicon; aluminum; tungsten; titanium; manganese; calcium; boron.

Table VI
Temperatures of Cathodes

Tube Class	Brightness Temperature — deg K		
	Minimum	Average	Maximum
Mercury-Vapor and Gas Rectifiers			
Heavy Output and Certain Acorn Tubes	1076	1100	1125
Power Amplifier Receiving Types — General			
Cathode-Ray Types	1050	1075	1100
Low Grid Current Types With Regular Grids			
High Power Output Receiving Types	1000	1025	1050
Low Power Output Receiving Types	975	1000	1025
With Special Grids	Temperature depends on grid design		
Small Receiving and Low Output Types — General	975	1025	1100

Carbon in Pure Nickel — N81

It is known that a small amount of carbon in pure nickel acts as a fairly good reducing agent for oxide-coated cathodes systems during tube processing and early life. The reaction of carbon with barium oxide yields the desired barium for activation of the coating together with carbon monoxide gas. The amount of carbon in this alloy is maintained below 0.02 per cent as an upper limit in order to prevent the excessive production of barium and the concurrent evaporation of barium that would give rise to grid emission.[14] However, this limitation on cathode activity results in a lower level of emission performance than is desired initially. Increasing the carbon content to 0.05 per cent will increase the production of barium to give improved emission performance but can contribute to grid-emission effects. It should be noted that another effect of excess carbon in the base metal (in excess of 0.05 per cent) is to introduce an extra amount of carbon monoxide gas into the tube which must be completely taken up by the getter. Too much carbon in the base metal can cause loss of tubes during production because of excessive gas.

Carburized N18 Cathodes

Cathodes made from a particular lot of N18 alloy have been processed by gas-carburization technique in order to increase the carbon content[15] and thus improve initial emission performance and obtain higher transconductance. The reason for this procedure was the poor performance of this particular N18 alloy. The cathodes were heated in an atmosphere of hydrogen and methane. The amount of carbon introduced onto the cathode surface by the process was about 0.05 per cent. Normally, the N18 alloy would contain about 0.02 per cent silicon, 0.025 per cent magnesium, and about 0.02

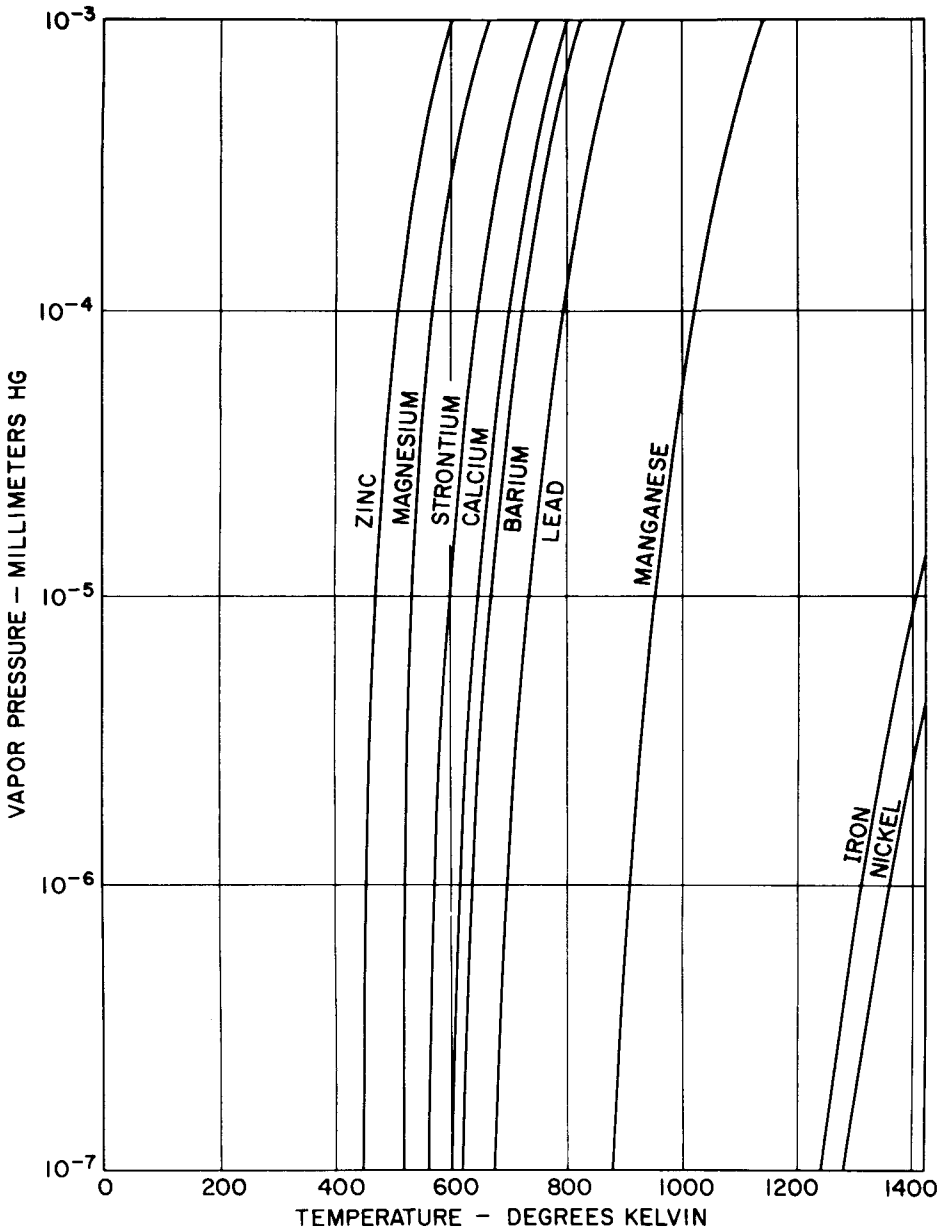


Figure 4. Vapor Pressure Curves for the Elements (by R. E. Honig[12])

per cent carbon after hydrogen firing. The tube type using these particular alloys was experiencing low initial emission. All analyses and investigations gave evidence that the reducing agents were combining as oxides or intermetallic compounds, or were being prevented from diffusing to the surface by oxide inclusions or carbides. The carburizing treatment added the extra amount of carbon needed to create activation prior to the point at which sufficient diffusion of the main reducing element content could supply the necessary cathode activity. There were no reports as to excess gas in the tubes.

Carbon and Magnesium

Magnesium and carbon can form carbides. It is known that MgC_2 is unstable at elevated temperatures; it decomposes to form carbon and the stable Mg_2C_3[16]. This intermetallic compound can form in vacuum-melted nickel when the carbon content is as high as 0.07 to 0.10 per cent. A cathode nickel alloy containing such high carbon content failed to sublime magnesium even though the latter element was present to the extent of 0.04 to 0.07 per cent. The standard ASTM test was negative; tubes made from this melt had low emission and were gassy. When the cathodes were baked in hydrogen (dew point 35 C) at 800 C for ten minutes, the carbon content was reduced to 0.03 per cent. Sublimation tests performed on these processed cathodes gave indication of the presence of magnesium on the target. Tubes made from these treated cathodes gave normal emission performance and were not gassy. It may be concluded that excessive carbon content in the nickel alloy can affect the reducing activity of magnesium and slow down the rate of diffusion and sublimation by forming carbide compounds.

Magnesium

Magnesium and silicon are the two elements that have received the most attention as reducing agents for oxide-coated cathodes. Both have been deemed desirable, even if not absolutely necessary. Magnesium is used because its presence insures high initial activity of the cathode with the result that good electron emission is readily obtained through the early life period; silicon is used because, although it acts more slowly, its presence maintains a high cathode activity over the long life period. Modern tube performances, however, require cathodes that give low interface impedances. As a result, silicon has been relegated to a secondary place and magnesium is now receiving first consideration in the design of cathode alloys. On the other hand, the very factors that make magnesium so desirable for easy activation of the oxide coating also give rise to the undesirable side effects of sublimation, electrical leakage between electrodes, grid emission and anode emission due to the copious amounts of barium produced in the emission coating and evaporated onto the adjoining electrodes. Considering the above factors, it can be concluded that the magnesium content should be held to a lower limit of 0.03 per cent for ease of activation of the coating and must not exceed 0.07 per cent to prevent the effects of excessive sublimation and excessive production of barium in the coating.

Silicon

Silicon retains its activity as a reducing agent over a period of thousands of hours because of its slow rate of diffusion. Its presence in cathode nickel alloys insures good emission performance over a long life period. However, it does have an interfering side reaction. It reacts with barium oxide to form an interface compound, barium orthosilicate, which has a dark grey color, and when present to the extent of 0.20 per cent introduces an electrical interface impedance of hundreds of ohms. The operating temperature of the oxide coating can be lowered by as much as 45 C because of the increase in radiation which results from the dark grey color of the interface compound. Under normal conditions, current flow keeps the interface compound in an active condition due to the presence of barium donor sites and, therefore, little resistance is offered to the flow of current. However, when the tube is operated under intermittent or standby conditions, the interface becomes deactivated and its impedance increases to a large value. This phenomenon usually occurs in tubes that are used for pulsed emission, computer circuitry, and especially with tubes that have high values of transconductance. When the silicon content in the nickel alloy is excessive, a high rate of production of barium and a subsequent increased rate of evaporation of barium causes grid emission to occur in some critical tube types. In general, experience has indicated that there is a greater tendency for peeling of coating to occur when the silicon content of the base metal is greater than 0.20 per cent. Whenever interface impedance is a factor that can influence the performance of the tube, the silicon content should not be greater than 0.02 per cent.

Aluminum

The results obtained in the laboratory and factory indicate that aluminum is a fairly active reducing agent and can be helpful in obtaining long life from oxide cathodes. However, the aluminum content in the nickel alloy should be limited to 0.10 per cent in order to minimize peeling of the coating that apparently occurs in conjunction with an excess of aluminum. At least, 0.05 per cent magnesium and 0.03 per cent silicon should be present when aluminum is part of the alloy composition in order to insure adequate activity and life.

Tungsten

Investigations have indicated that a nickel cathode alloy having a low silicon content with 4.7 per cent tungsten does not exhibit measurable interface impedance when cathodes made from it are life tested under conditions which would produce high impedances with cathodes made from high silicon nickel cathode alloys.[17] X-ray diffraction studies of the interface, by Eisenstein, failed to reveal the barium tungstate compound upon completion of a 6000-hour life test. Barium orthosilicate was found to be present as the interface compound. Analysis of the cathode base metal indicated that the composition included 0.02 per cent silicon as well as the 4.7 per cent tungsten. It is concluded that the silicon reacted with the barium tungstate to release the tungsten and to form the barium orthosilicate compound. The orthosilicate compound was maintained in a more active condition by the further reaction of the tungsten with the barium oxide to provide the activating barium donor centers.

Some evidence has been obtained to indicate that a nickel alloy containing 4 per cent tungsten plus 0.04 per cent magnesium and containing no silicon had caused very little decay in emission on life. The measured interface impedances were less than 5 ohms/cm^2 at the end of thousands of hours.

The chemical reactions of tungsten-bearing nickel alloys with barium oxide are well described by Hughes et al.[18] At best, a tungsten nickel alloy has only a mild reducing activity; it provides enough barium to yield good activation of the oxide coating but it is not active enough to withstand the depletion effects of ordinary tube poisoning. Low emission and extreme slumping during early life is experienced, whereas nickel alloys containing the same amount of tungsten but with small percentages of magnesium and silicon can withstand the same degree of poisoning effects. In the absence of such poisoning effects, tungsten nickel cathodes can give very long life performances. It is concluded that the tungsten content of the nickel alloys should be held to a maximum of 4 per cent for purposes of imparting hot strength to the alloy and contributing some reducing element activity. In addition to this tungsten content, the cathode alloy must contain the more active reducing agent such as magnesium and silicon to initiate good production of barium in the coating. The operating temperature of the oxide coating can be lowered by as much as 50 C, because of the increase in radiation which results from the dark-brown color of the barium tungstate interface compound.

Titanium

When all the information available on the use of ti-

tanium as a reducing element agent in the cathode nickel alloy is considered, it is evident that titanium, although it is a powerful reducing agent and imparts excellent hot strength to the alloy even when used alone, should not be used in amounts greater than 0.10 per cent because a dark-colored interface compound, barium titanate, is formed. The decrease in cathode temperature due to the increase in radiation, the increase in interface impedance due to the compound formation, and the tendency for the coating to peel when titanium is present in substantial amounts are reasons for not using titanium as a prime activating agent. Cathode alloys should not contain more than 0.005 per cent titanium, the minimum amount detectable by spectrographic techniques.

Manganese

Manganese is considered to be a mild reducing agent. It is added to nickel alloys for purposes of cleaning the metal and as an aid to the hot forging of the ingots. The specifications for air-melt grades of nickel alloys allow a maximum limit of 0.20 per cent manganese to be present in the alloy. Vacuum-melt grades of nickel alloy permit a maximum of 0.08 per cent manganese to be present in the alloy.

The effects of adding manganese as a reducing agent to pure nickel has been investigated;[19] no effect on improvement of emission performance has been found. It was evident that the presence of more than 0.20 per cent of manganese could induce side effects that lowered the transconductance. No effect on the contact potential between grid and cathode has been ascribed to the presence of manganese. Sublimation studies have indicated that manganese is present in the deposits found on the targets when the cathodes contained more than 0.15 per cent of manganese. Some evidence exists that manganese may inhibit the diffusion of magnesium in nickel at 850 C but the use of multicomponent alloys complicates the interactions. Increased stabilization or increased rates of diffusion and evaporation can be produced on any one of the components by interaction of the other agents that may be present.

In general, manganese is considered to be an undesirable additive for cathode base metal purposes. Manganese content is limited to a maximum of 0.08 per cent and should be maintained at 0.05 per cent in those cathode alloys that are to be used in critical tube types because there are measurable leakage effects associated with the presence of too much manganese.

Calcium

The presence of elemental calcium has been detected in amounts approaching 0.001 per cent. Calcium is obtained from the refractory lining used for crucibles wherein the nickel alloy is melted. The reducing activity of carbon is sufficient to convert some of the calcium silicate content of the lining to elemental calcium. The presence of calcium in excess of critical amounts apparently causes embrittlement of the nickel alloy. Its presence in the interface region apparently causes poor coating adherence and high impedance effects. Further studies are necessary to substantiate these interactions.

Boron

Commercial processes for the manufacture of nickel alloys use small amounts of boron to improve the working qualities of the metal. Analyses indicate that it is present in the nickel alloys to the extent of 0.001 per cent. It is considered that where boron accumulates in harmful concentrations at the surface of the cathode, it can form a high-impedance barium borate glass which offers such a restrictive surface barrier to the diffusion of other elements that poor emission performance is obtained. All precautions should be taken to keep the boron content at or below 0.0005 per cent. It is especially important that boron be absent from the lubricating fluid used in drawing and fabricating operations in order that the surface layers of the nickel are not contaminated with boron.

CHARACTERISTICS OF SELECTED CATHODE ALLOYS

Temperature ranges of oxide-coated cathode systems for tube design purposes are given in Table VI. A cathode material selection guide as given in Table VII and a list of the limiting compositions of domestic cathode nickel alloys as tabulated in Table II give the characteristics of cathode nickels commonly used by RCA in the manufacture of electron tubes. Dr. I. Stacy has compiled a guide to cathode nickels to acquaint the tube design engineer with the properties of the new developmental air-melt alloys submitted for testing by commercial producers and the more recent vacuum-melt alloys made by RCA for use in computer and industrial tube types.[20]

The relative activity of a cathode nickel alloy refers to its capability for producing barium by the reaction of the reducing element content of the metal with the barium oxide component of the emission coating. The activity of an alloy is related to the effective concentration of the reducing agent (i.e., not combined as an oxide or an intermetallic compound), the rate of diffusion of the reducing agent through the alloy, and the extent of surface film formation that can impede the diffusion of the reducing element into the emission coating. RCA vacuum-melted alloys are so well-controlled as to composition and effective concentration (minimum combined compounds) that little or no difference exists in the activity of melts of the same composition. The choice of composition and activity of a cathode alloy must consider the particular tube application in terms of temperature of processing and operation and the sublimation effects of leakage, grid emission, and contact potential as well as the growth of the interface impedance over the life period.

REFERENCES

1. Sidgwick, N. V., The Chemical Elements and Their Compounds: Group VIII, pp. 1316-1620, Clarendon Press, Oxford, 1950
2. Mudge, W. A., "The Melting of Nickel," AIME Symposium on Melting Non-Ferrous Metals, 1946
3. Merica, P.D., and R.G. Waltenberg, "The Nickel Industry," Technical paper, Bur. of Stds. No. 281, 1925

Table VII

Cathode Selection Guide

Manufacturer's Designation*				Form	Ease of Activation	Interface	Sublimation	Leakage	Strength	Peeling	Reducing Activity
RCA	INCO	ST	DH								
N18	220	G11	---	seamless cathodes	average	medium to to high	medium	medium	average	often	normal
N34	225	G3	---	strip	easy	high	medium	high	average	often	active
N42	---	---	N42	strip	easy	high	medium	medium	average	rarely	active
N80	---	G6	399	strip	easy	high	slight	slight	average	some	active
N81	---	G21	499	strip	difficult	low	none	slight	weak	none	passive
N109	330	---	---	seamless cathodes and strip	average	medium	medium	high	average	rarely	normal
N120	---	G5	699	seamless cathodes	average	medium	medium	medium	average	some	normal
N131	---	P50	999	seamless and lockseam cathodes	difficult	low	none	slight	weak	none	passive
N130	---	A31	---	seamless and lockseam cathodes	difficult to average	medium	slight	medium	strong	rarely	normal
N132	---	---	---	seamless cathodes and strip	difficult to average	medium	slight	slight	average	no data	normal
N133	---	---	---	seamless and lockseam cathodes	difficult to average	medium	slight	medium	strong	rarely	normal
N136	---	---	---	seamless cathode and strip	average	medium	medium	medium	average	rarely	normal
N144	---	---	---	seamless and lockseam cathodes	no data	no data	no data	no data	no data	some	passive
N702	---	---	---	seamless cathodes	no data	no data	no data	no data	very strong	no data	no data

*RCA Radio Corporation of America; INCO = Int'l. Nickel Co.; ST = Superior Tube Co.; DH = Driver-Harris

4. Code S1
5. Brick, R. M., and H. Philips, The Structure and Properties of Alloys, Mc-Graw-Hill, New York, 1942
6. Wise, E.M., and R.H. Schaefer, The Properties of Pure Nickel Metals and Alloys, p. 1067, December, 1942
7. White, A. D., "Application of Thermodynamics to Chemical Problems Involving Oxide Cathodes," Jour. Appl. Phys., Vol. 20, p. 856, 1949
8. Rittner, E.S., "Theoretical Study of the Chemistry of Oxide Cathodes," Philips Research Rpt., Vol. 8, p. 184, 1953
9. Peterson, R. W., D. E. Anderson, and W. G. Shepherd, "Influence of the Cathode Base on the Chemical Activation of Oxide Cathodes," Jour. Appl. Phys., Vol. 28, pp. 22-31, 1956
10. Code S2
11. ASTM Specifications F-278-52T, American Society for Testing Materials
12. Honig, R. E., "The Vapor Pressure Curves for the More Common Elements," RCA Review, Vol. 18, p. 195, 1957
13. Code S3
14. Code S4
15. Code S5
16. Rueggeberg, W.H., "The Carbides of Magnesium," Jour. Amer. Chem. Soc., Vol. 65, p. 602, 1942
17. Eisenstein, A., H. John, and H. Affleck, "Double Activators in Oxide Cathode Base Alloys," Jour. Appl. Phys., Vol. 24, p. 631, 1953
18. Hughes, R.C., P. P. Coppola, and H. T. Evans, "Chemical Reactions of Barium Oxide on Tungsten Emitters," Jour. Appl. Phys., Vol. 23, p. 635, 1952
19. Code S6
20. Code S7

The Interface and Other Oxide-Cathode Impedances

E. J. Hannig

Harrison

Although the oxide cathode was discovered by Wehnelt[1] in 1904, it was sixteen years later that the problem of cathode impedance was recognized by Arnold[2] who observed a platinate interface layer on platinum-iridium filaments in 1920. In that year, few circuits demanded anything more than "some" emission. Practically any tube would be sufficient to meet circuit demands. The period between the '20s and the '50s was one of rapid advancement for the entire electronics industry. Wide-band amplifiers, large- and small-signal pulse circuits, radars, computers, and the like all demanded more than just "some" electrons from cathodes. Demands for tubes capable of 10,000 to 20,000 hours of reliable service only placed heavier burdens on the oxide cathode. It was during this era that the electronics industry became aware of the cathode impedance in the more critical applications when the cathode had to have practically zero impedance for as long as possible. In this period, Eisenstein[3] published a comprehensive paper on the common cathode-interface layer, barium orthosilicate, and Nergaard and Matheson[4] summarized the role of the interface impedance in electron-tube circuits. In the period from 1944 to 1954, oxide-cathode literature dwelt at great length on cathode impedance.[5-10] It is the purpose of this article to analyze cathode impedances, to establish some methods to minimize them, and to review some of the more acceptable means for measuring them.

CATHODE SLEEVE ACTIVITY

Most cathode impedances are influenced by the characteristics of the cathode sleeve material. Although nickel is commonly used, certain alloys of nickel are acceptable for fabrication into cathode sleeves and others are not.[11] It is usual to find tube types that can be processed satisfactorily only if a certain nickel alloy is used as the cathode sleeve material. Even different melts of the same alloy may require different processing schedules. These are only a few of the variables which tube processing must accommodate and still produce the maximum quantity of acceptable tubes in minimum time with minimum shrinkage.

The three general classifications of nickel alloys are given in Table I, together with the reducing-agent content of each. The classifications describe the speed with which these alloys can activate an oxide coating. For example, alloy N34 will activate the oxide much more readily than any of the others listed in Table I. Consequently, the use of this material in a vacuum tube requires less time and power in processing; it is called an active alloy. Alloy N81 is the most difficult to process. For high-speed production, an active alloy is most desirable although, unfortunately, the more active alloy is often prone to application failures.

Table I

Classification	Alloy No.*	Maximum Reducing-Agent Content--Per Cent (Weight)				
		Mg	Si	Mn	Ti	C
Passive	N81	0.005	0.010	0.02	0.005	0.02
Normal	N109	0.03-0.10	0.01-0.05	0.20	0.005	0.08
Active	N34	0.01-0.08	0.15-0.25	0.20	0.03	0.04
Normal	N18	0.01-0.10	0.01-0.05	0.20	0.03	0.04

*The alloy numbers are RCA designations for commercial materials. A more complete listing of these alloys is given in Tables II and VII in the article on "The Nickel Base Metal for the Oxide Cathode."

Cathode base-metal sublimation testing has been standardized[12] by the American Society for Testing Materials for the more precise evaluation of alloy activity. In the test, the resistance between two conducting strips painted across a sheet of mica is continuously measured. The bare nickel sleeve is mounted opposite these strips. The sleeve is heated by a conventional cathode heater to some given temperature; a brightness temperature of 900 C has been satisfactory for most alloys. Materials from the hot nickel sleeve will sublime and deposit a thin film on the mica causing the resistance between the conducting strips to decrease at a rate which depends principally on the amount of metal coming out of the sleeve and the rate at which it comes out.[13] These metals are primarily magnesium, manganese, and nickel. Characteristic sublimation curves for alloys N18, N34, N109, and N81 are shown in Fig. 1. Less than one hour is generally required to obtain such curves. The curves give a good indication of the availability of magnesium in the cathode during processing and activation. Magnesium and manganese appear to be the primary activating agents. However, at some point in life, other, more long-lived agents such as silicon take over the function of producing barium for emission.

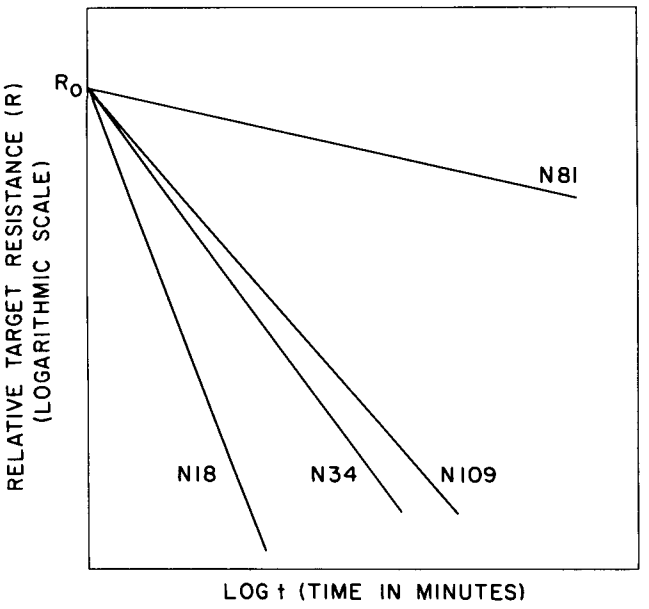


Figure 1. Graph of Target Resistance Plotted Against Elapsed Time for Different Alloys (R_0 is the target resistance at t = 1 minute; resistance units are arbitrary)

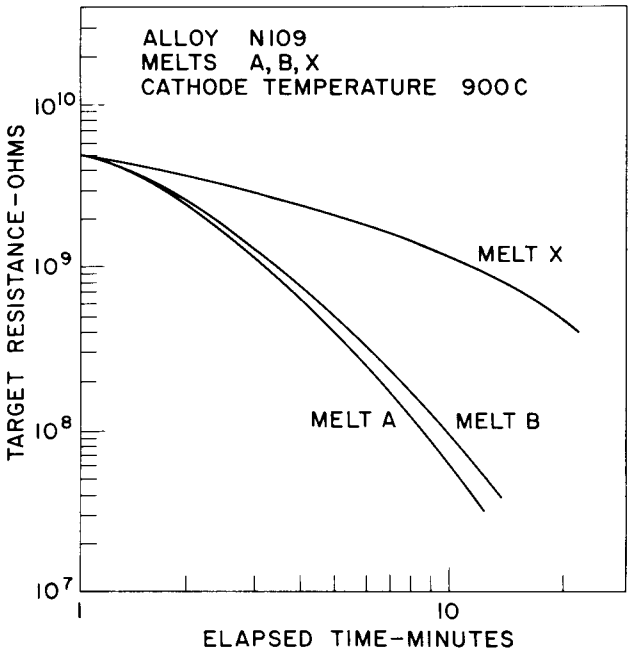


Figure 2. Sublimation Curves for Three Melts of the Same Cathode Base-Metal Alloy

The electron-tube industry requires a wide range of cathode nickel alloys to make the many tube types required. However, only slight variations in the characteristics of an alloy can be tolerated from melt to melt. Each melt must meet certain compositional specifications or it is rejected before it ever reaches the tube. Unfortunately, melts with apparently identical compositions may have widely different sublimation and activation characteristics.

Consider a factory problem of some years ago. Alloy N109 cathode material was being used in the nickel sleeve for a tube type. No particular difficulties had come from this type for many months; many melts of N109 had been used during production with only slight variations in processing. Then, melt X made from alloy N109 was tried with chaotic results. It was impossible to properly activate oxide cathodes using melt X as the cathode sleeve. The factory tried melts A and B and produced good tubes. All three melts were quickly rechecked, both by wet analysis and by spectrographic analysis, and all were within specifications. Other standard tests were made with little success. So far as all these tests were concerned, the three melts were interchangeable. Cathode sublimation testing, then in its infancy, was attempted. The results of this test (Fig. 2) were conclusive.

From the sublimation curves it was at once apparent that melt X did not belong in the same activity category as melts A and B, even though all other analyses indicated that it was. Some condition within melt X was inhibiting the free sublimation of its reducing agents. Melt X, therefore, produced cathodes very much underactivated, and processing and aging conditions could not be varied sufficiently to produce the same oxide cathode as that activated by melts A and B.

While melt variations such as the above do not happen every day, they occur often enough to warrant mention. Often, the presence of bulk impedance (the impedance of the oxide coating excluding any interface impedance) in the oxide coating is the result of such variations. The cause of all cathode impedances can generally be traced to the base metal; the amount and availability of its reducing agents determine whether or not a cathode impedance is likely to occur.

CATHODE ACTIVATION

This section will discuss only enough of the cathode activation process to aid in understanding cathode impedance. Although it is common commercial practice to employ the triple carbonate as cathode spray, it is only necessary to refer to the simple barium carbonate ($BaCO_3$) as the material with which the cathode is coated initially.

During the processing of an oxide-coated cathode, the following reaction always precedes activation:

$$BaCO_3 \xrightarrow{800\ C} BaO + CO_2 \uparrow \qquad (1)$$

The carbon dioxide (CO_2) is pumped off and the carbonate coating is converted to barium oxide (BaO). The cathode temperature is then raised to approximately 1000 C or higher; this temperature favors the more rapid diffusion of the reducing agents from the base metal into the oxide coating. The reaction that occurs at the cathode operating temperature is described by Eqs. (2) and (2a):

$$BaO + Mg \rightarrow Ba + MgO \qquad (2)$$

$$2BaO + Si \rightarrow 2Ba + SiO_2 \qquad (2a)$$

In the case of initial activation, these reactions occur

at about 1000 C; normal tube operating conditions require only approximately 800 C. In Eqs. (2) and (2a), magnesium and silicon are taken as the reducing agents since these can be considered as the prime reducers, although there are others of less importance contained in the base metal (see Table V in the article on "The Nickel Base Metal for the Oxide Cathode").

The diffusion constants are given in this same Table V and are related to cathode temperature as $D = f(e^{-1/RT})$. The outward flux $D(dc/dx)_1$ of activator particles is given by Rittner[14] as:

$$\frac{2D\,C_o}{l}\,e^{-t/\tau} \qquad (3)$$

where D = diffusion coefficient
C_o = initial concentration of activator in the base metal
l = thickness of base metal (approx.)
t = time in which diffusion takes place, and
τ is a constant $= \frac{4l^2}{\pi^2 D}$

Cathode activity depends principally upon the temperature and $D(dc/dx)_1$. The dependence on outward flux is a property of the base metal; it distinguishes one nickel alloy from all other nickel alloys. As a result, cathode temperature control is a significant factor in cathode activation; it determines the time required to reach the desired cathode activity.

As the operating time of a cathode increases to thousands of hours, $e^{-t/\tau}$ in Eq. (3) becomes important in that the activator is gradually being depleted. Therefore, the flux decreases with life. The rate at which reactions (2) and (2a) proceed decreases due to a decrease in activator flux. Consequently, the cathode becomes deficient in the barium required to maintain a good conductive coating. Emission may or may not suffer, although it usually follows the increase in cathode resistance on long life. Cathode environment plays a more substantial part in causing low emission problems; any condition causing excessive evaporation of surface barium will decrease the emission more readily.

CATHODE IMPEDANCES

If reactions of the type described in Eqs. (2) and (2a) can be completed during the time allowed for processing, the cathode will emerge as a good emitter that is relatively "cathode-impedance-free." If the reactions are retarded in any way, cathode bulk impedance will result. This condition will require aging or some period of operation to eliminate the impedance and produce the good conductive coating. If the rate of this reaction is ever insufficient to maintain good coating conductivity, bulk impedance results again; should bulk impedance develop after many thousands of hours of operation, the possibilities of recovery are extremely slight.

While the reduction reactions Eqs. (2) and (2a) are taking place, a third, extremely deleterious, reaction is taking place as a by-product of Eq. (2a):

$$2BaO + SiO_2 \longrightarrow 2BaO\cdot SiO_2 \qquad (4)$$

Silicon dioxide, which is formed together with free barium in the chemical activation equation, reacts with the oxide of barium to form barium orthosilicate. This compound exists at the interface of the base metal and the oxide coating. Its properties approach that of an insulator; its electrical conductivity is much lower than that of any other possible cathode compound, which may include tungstites and titanates among others. These two are objectionable interfaces; however, tungstites are possible only with certain more recent alloys, and titanium is held to an extremely low value in practically all alloys. Their conductivity is, nevertheless, low enough to cause considerable concern should they be present in sufficient volume. Thus far, the best way to inhibit interface impedance is to limit the silicon content of cathode alloys to small values. In this way, C_o in Eq. (3) has a small value, consequently, the flux density of the particular reducing agent is decreased.

Because the orthosilicate layer has such high resistance, one must ask, "How, then, is it possible for cathode current to flow at all?" The best answer to this question at the present time lies in the realm of semiconductor theory. It is analyzed by Nergaard and Matheson[15] in terms of an impurity-type semiconductor having the same donors as that of the barium oxide.

A third cathode condition which manifests itself as an impedance arises from poor coating adherence. Poor bonding of coating to base metal results in a peeling cathode. Such a cathode is more likely to occur with highly active alloys than with others. Sometimes the interface ($2BaO\cdot SiO_2$) is thick enough to physically separate portions of the coating from the cathode. Peeling cathode trouble is inherent in random melts of an alloy. The entire coating may peel on only limited areas. Thus far, it has been difficult to predict whether a given cathode would or would not peel; in this respect, peeling differs from the actual interface compound. It has relatively the same effects as the interface and bulk impedances. Interface-impedance investigations naturally led to peeling-cathode investigations. Instruments employed to evaluate the interface also measure impedance due to poor coating adherence.

Most cathode problems are the result of cathode environment which tends to suppress or accelerate the activation reactions. These reactions give rise not only to the common cathode problems,[16] but also to the cathode impedances, all of which limit the useful life of a cathode.

If the more obvious cathode ailments are excluded, it is possible to view the cathode as being in one or more of the four possible conditions described in Fig. 3. Quadrant I shows a cathode coating having an abundance of barium both on the surface and within the coating; this abundance permits good emission and virtually perfect conduction through the coating. Quadrant II illustrates the possible effect of bulk impedance caused by a smaller amount of barium within the coating and possibly on the surface; consequently, emission may be satisfactory but the coating itself is far more resistive than that of Quadrant I. Quadrant III shows the presence of the interface compound between coating and base metal, and a slight decline of barium in the coating, and

points out some barium within the interface itself; again, the result is a very poorly conducting cathode. Quadrant IV approximates the situation in a peeling cathode; some areas are as good as Quadrant I for conduction, whereas others are either loosely bound or have no coating whatsoever.

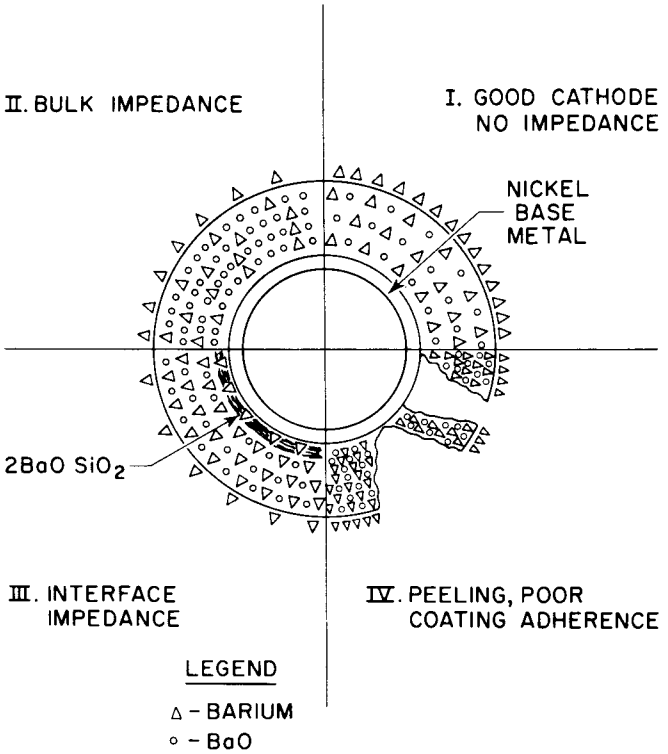


Figure 3. Cathode-Condition Diagram

EQUIVALENT CIRCUIT OF CATHODE IMPEDANCES

To study the effects of a cathode impedance, it is necessary to assume some sort of configuration for the passive components of the impedance. Cathode impedance can be approximated by the oversimplification in Fig. 4A; another, somewhat closer, approximation is given in Fig. 4B (see Ref. 15). Although neither representation accounts for every case of tube failure due to cathode impedance, Fig. 4A is adequate as a working model to demonstrate principles. The time constant, which is the product of resistance and capacitance, distinguishes one cathode impedance from another in nondestructive testing.

$$t = RC \qquad (5)$$

where t = time constant in microseconds,
R = resistance in ohms, and
C = capacitance in microfarads.

With regard to the direct-current effects which these impedances cause in a tube, it is obvious that some increase in bias can be expected, provided that the cathode current does not drop too radically. In either event, the direct-current parameters of the circuit are altered and the tube can be a failure for this alone.

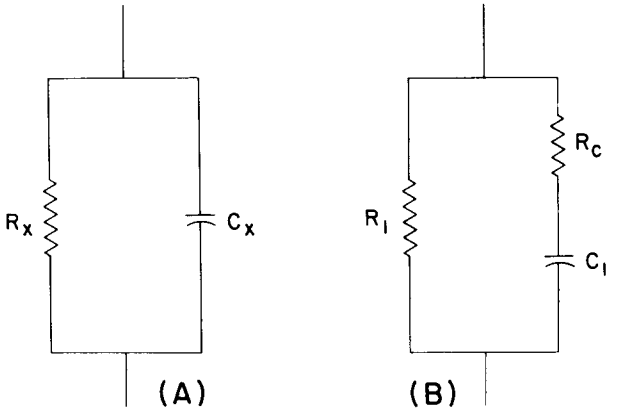


Figure 4. Simplified Equivalent Circuits of Cathode Impedance: (A) extreme simplification; (B) closer approximation

The effect of added resistance in the cathode circuit is illustrated in Fig. 5. The tube was first set to a normal operating condition of 4.4 milliamperes per volt (4400 micromhos). Then a 120-ohm resistor was added in series with the cathode as shown in the lower right quadrant of Fig. 5. The result was a bias shift of -0.3 volts, plate current was reduced by 15 per cent, and the mutual conductance g_m fell to 2.80 milliamperes per volt (2800 micromhos). The effect on the amplitude of the plate signal is shown in the upper right quadrant. Transconductance degenerated according to the formula:[4]

$$g_m \approx \frac{g_{mo}}{1 + g_{mo} R_k} \qquad (6)$$

where g_{mo} = initial g_m of stage,
R_k = any added cathode resistance,
g_m = g_m of tube with cathode resistance added.

Eq. (6) implies that high-transconductance tube types are more sensitive to small changes in cathode-resistor value than lower-transconductance types.

The resistance added to the cathode circuit simulated one component of cathode impedance. However, this resistance is not semiconductive and it is not temperature-dependent. A capacitance (C_k) added in parallel with R_k (Fig. 6) makes the shift in g_m a function of the time constant. The shift in g_m varies inversely with RC; the smaller RC becomes, the more closely g_m approaches the maximum shift permitted by R_k alone. Therefore, smaller cathode time constants are always more damaging to circuit performance.

Raudorf[17] has demonstrated (Fig. 7) that the gain of a circuit depends upon the frequency in a tube where cathode interface impedance exists; this characteristic is typical of short time-constant cathode impedances. Fig. 7 illustrates the frequency-dependent response of a cathode which has developed an interface impedance. The impedance produced by peeling cathodes would present about the same gain-shift pattern with one exception; the gain would approach its design value at a frequency mugh higher than that shown in Fig. 7. Its time

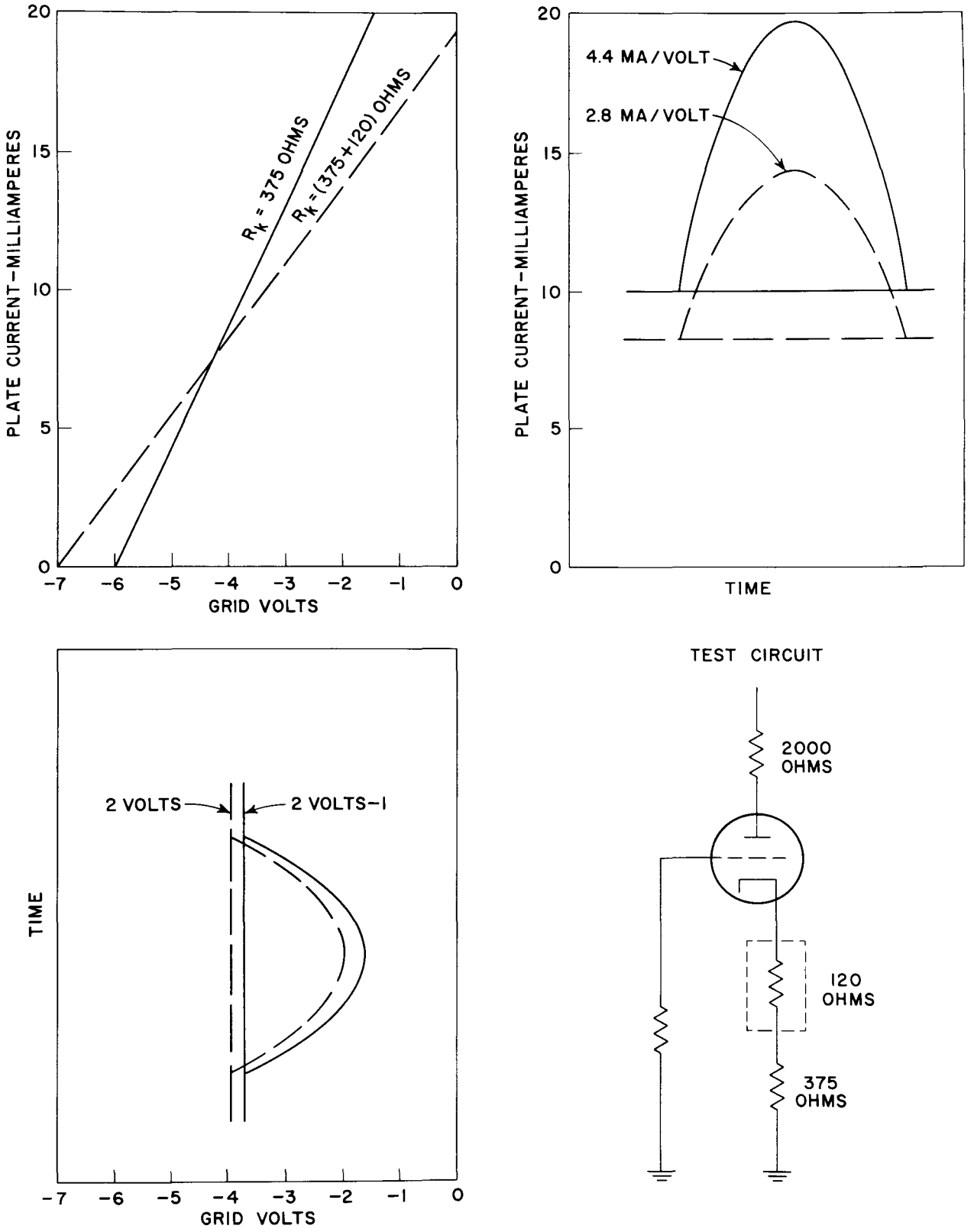


Figure 5. The Effect of a Bias Shift of -0.3 Volt on the g_m Characteristic of a 4.4 ma/volt Tube in an Actual Circuit (this effect was produced by physically placing a 120-ohm resistor in the cathode circuit)

constant is anywhere from 0.01 to 0.1 times that of the interface. The bulk impedance would allow a lower frequency response because the bulk time constant can be from 10 to 100 times that of the interface. The most significant effect on alternating-current parameters produced by these impedances is a frequency-

dependent reproduction of the input signal.

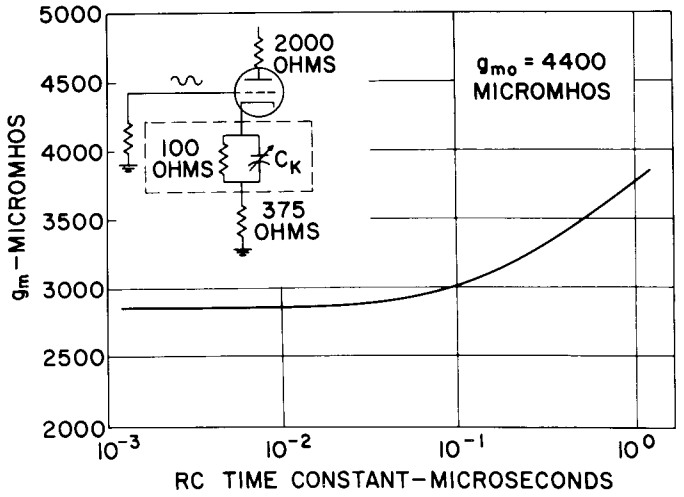


Figure 6. Curve Illustrating the Effect of Cathode Impedances of Different Time Constants on Circuit Response to a 10 Kc Signal (the design g_m (or g_{mo}) is 4400 micromhos; R_k is 120 ohms in all cases)

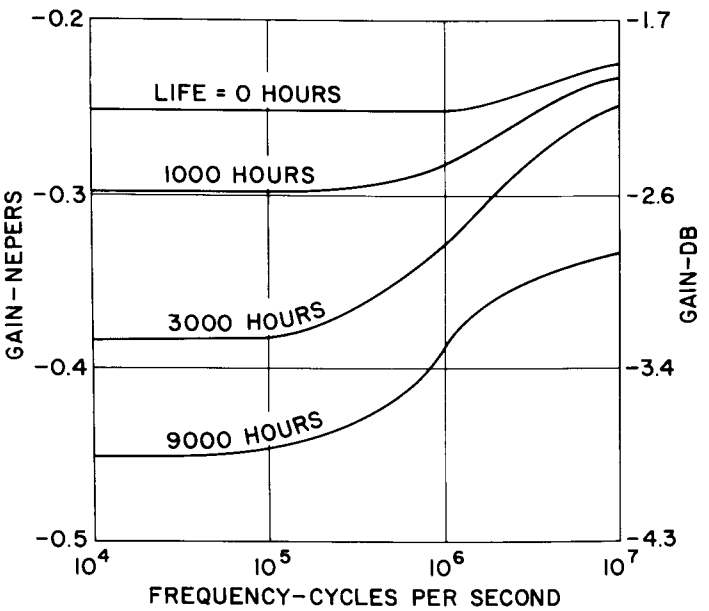


Figure 7. Variation of Gain with Frequency Caused by the Formation of Cathode Interface Impedance (the increased gain-shift with life is caused by increased interface resistance (R_i)

DECAY CHARACTERISTICS

Pulse-emission testing has disclosed several decays characteristic of oxide-coated cathodes.[18,19] Although many of these emission decays are due to deterioration of the emission mechanism, the interface decay and, possibly, the bulk decay has been mentioned in the literature. Poor adherence of the coating should produce the same effect as the others. There is currently very little in the literature on decay caused by poor adherence. Perhaps the extremely short time constants of peeling cathodes have hitherto been undetectable due to the longer pulse rise time on pulse-emission equipment. Many tube failures in computer and other pulse applications are traceable to one or more of these impedances.

Most pulse circuits require reasonably good fidelity, i.e., the output signal waveform must be similar to the input signal waveform. A cathode impedance causes the current waveform to deviate from the grid pulse voltage in a manner somewhat analagous to that of a capacitive discharge. For the first few hundredths of a microsecond, the cathode current follows the grid voltage; thereafter, the blocking action of the impedance increases exponentially to produce a corresponding decrease in the current, as shown in Fig. 8B and 8C. The amount of this degeneration and the resultant waveshape depend upon the resistance and time constant of the impedance.

Should a cathode impedance develop at some intermediate stage of an amplifier, only signals of the types shown in B and C of Fig. 8 would be available to the succeeding stages, and a distorted waveshape would appear at the output of such an amplifier. Circuits requiring definite values of triggering voltage would fail upon receiving the pulse as modified by the cathode impedance.

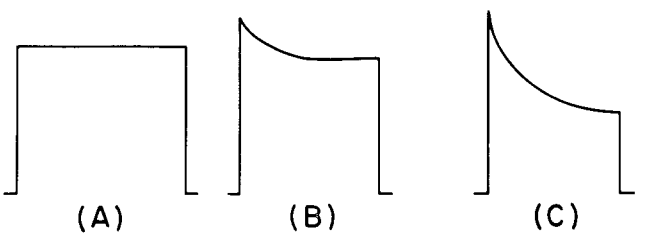


Figure 8. Current Decay Produced by Cathode Impedance: (A) the input voltage; (B) the distortion produced as a result of cathode impedance with a short time constant; (C) the distortion produced by a cathode impedance with a longer time constant than in (B)

The electrical effects of the cathode impedance are nearly always manifested as a direct-current bias shift plus a frequency-dependent degeneration in the affected stage. These two effects adequately describe both alternating-current and pulse response. The range of the degeneration depends upon the type of impedance. These impedances can be categorized by their time constants. The bulk impedance, or oxide impedance, will have a time constant greater than 10 microseconds; usually it will be on the order of 100 microseconds. Interface impedance will be about 0.5 to 5.0 microseconds.[20] Poor adherence of coating generally results in a 0.05- to 0.5-microsecond time constant. There is some overlapping in the effects produced by the three types of impedance.

EFFECT OF ALLOY ON CATHODE IMPEDANCE

Certain cathode alloys are prone to definite types of cathode impedance. Low-activity base metals (N81* and N131*) are often subject to bulk impedance. This effect is generally caused by under-processing the tube; it is often found after thousands of hours of operation as the result of loss of barium in the coating. Bulk impedance is a curable malady, as was shown by life test data on ten tubes (type 6AL5) which used N131 cathode material. Initially, the value of bulk resistance was between 0 and 105 ohms (average, 44 ohms). The tubes were put on conduction life and at the end of 500 hours the resistance of the worst tube was 5 ohms, and the average resistance was 0.4 ohm. The data of these tests are plotted in Fig. 9. The tubes had been under-processed initially but, with life, all bulk resistance values fell to a point where tube operation was satisfactory.

Alloy N34 is probably the most infamous of the interface-impedance-producing cathode materials. Its high silicon content (0.15 to 0.25 per cent) enhances the formation of large amounts of barium orthosilicate. N130 alloy, having approximately 4 per cent tungsten, gives rise to barium tungstate ($3BaO \cdot WO_3$). The author knows of no satisfactory cure for this condition other than careful selection of cathode alloys with strong emphasis on keeping the titanium, aluminum, and silicon content low. A temporary palliative for interface impedance is to draw a large emission current from the cathode. This large current temporarily activates the interface layer. However, when the tube is subsequently brought to its normal operating condition, the deleterious effect of the interface impedance will slowly return.

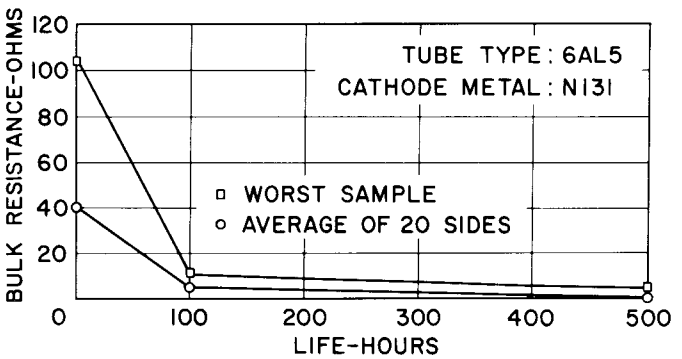


Figure 9. Variation of Cathode Bulk Resistance with Life

OPERATIONAL CAUSES OF INTERFACE IMPEDANCE

In addition to the interface-producing elements in a cathode, at least three other factors favor the growth of interface impedances.

These factors are:

1. Duty cycle. If an application requires very low current or allows the tube to stand in a nonconducting state for appreciable lengths of time, it fosters the growth of interface impedance. Low duty cycles promote rapid development of interface.

2. Cathode temperature. High cathode temperatures promote interface formations. High-silicon cathode nickels can have an interface immediately after processing if the cathode is run at too high a temperature.

3. Cathode current. The value of cathode current drawn in the application strongly influences the rate of formation of the interface. Lower currents produce high interface impedance in less time than do higher currents.

The effect of duty cycle has been demonstrated by Waymouth.[6] The results of his experiments are described in Fig. 10 where the percentage of survivors is plotted against hours of life. The tubes whose characteristics are indicated by the curve marked Panels I and II, drew 60 milliamperes per square centimeter during life for a 100 per cent duty cycle; the tubes whose characteristics are indicated by the curve marked Panels II and IV operated at cutoff, or zero per cent duty cycle. In this graph, tubes were considered failures as soon as the interface was detectable. The cathode alloy for all these tubes contained approximately 0.03 per cent silicon by spectrochemical analysis.

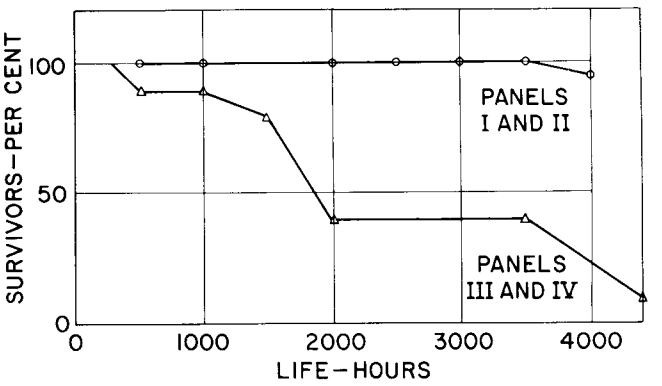


Figure 10. Effect of Duty Cycle on Interface Resistance

Cathode temperature has been shown by Metson[21] to bear directly on interface growth. Fig. 11 shows the result of running tubes for 3000 hours at different heater voltages (cathode temperatures). The mean value of the interface impedance of large lots of tubes is plotted against life time in hours.

The actual silicon content of the cathode sleeve was 0.05 per cent; were a base metal similar to N34 alloy used, the magnitude of interface developed would have been ten to a hundred times greater than that indicated.

A more quantitative evaluation of the role of silicon content in the cathode sleeve was obtained from a 100-hour life test. Twenty twin-triode tubes, type 6SN7GT, were run at a heater voltage of 7.5 volts; ten used N109 alloy with 0.01 to 0.05 per cent silicon, and ten used N34 alloy with 0.15 to 0.25 per cent silicon. The results are shown in Table II.

During this short life test, the tube was not permitted to draw current. The data illustrate the dependence of

* See the footnote at the bottom of Table I

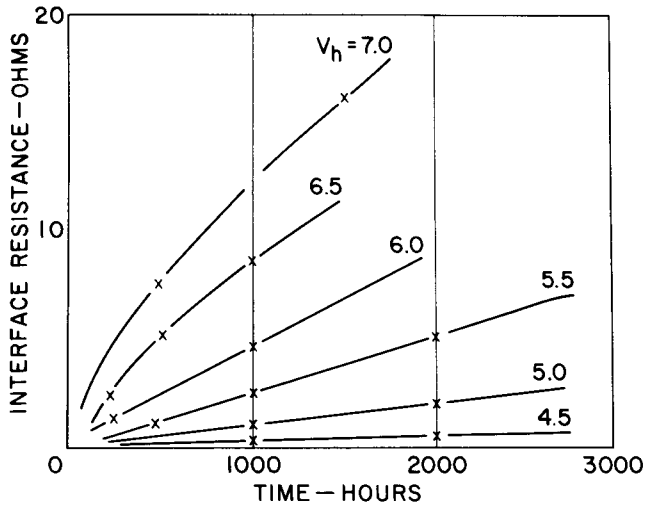


Figure 11. Rate of Growth of Interface Resistance as a Function of Temperature

interface growth on the supply of silicon in the base material.

of peeling on direct-current tube characteristics are slight and, like the bulk impedance, it has very little effect on cathode temperature. Once peeling starts on a cathode, there is no cure; the magnitude of the defect only increases with life. In addition to N109 alloy, the effect has also been detected on N18 alloy and N130 alloy.

Poor adherence of the coating can occur at the ends of the coated area, in a single portion of the cathode, or over the entire coated area. It can be caused by the base metal itself, by poor cleaning methods, or by large interface deposits which push the coating away from the nickel sleeve. Life data on peeling are not so copious as data on bulk and interface because of the difficulties in making coatings peel.

CATHODE IMPEDANCE EVALUATION

Methods for evaluating the cathode impedance generally involve either alternating-current or pulse techniques. Some investigators have used direct-current methods[22] but these methods have not met with acceptance in the tube industry. It is common practice to measure the cathode resistance (R_k) and the associated capacitance (C_k) or the time constant of the impedance.

Table II

N109 Tube No.	Interface Resistance—ohms		N34 Tube No.	Interface Resistance—ohms	
	Unit No. 1	Unit No. 2		Unit No. 1	Unit No. 2
1	0	0	1	48	38
2	0	0	2	29	38
3	0	0	3	45	50
4	0	0	4	50	94
5	0	0	5	66	50
6	0	0	6	41	30
7	0	0	7	57	75
8	0	0	8	90	69
9	0	0	9	80	83
10	0	0	10	50	75

COATING ADHERENCE

Because alloy N109 appears to be so much better than N34 for retarding interface formation, it has become widely used in computer and other critical tube types. As a consequence of this shift, the problem of poor coating adherence has become more prominent. Although alloy N109 is not a consistent peeler, many of its melts are prone to peeling of coating. The behavior of a peeling cathode is similar to the behavior of interface impedance, with the exception that peeling cathodes show a time constant between 0.05 and 0.5 microsecond. Beyond one microsecond, a pulse signal would be reproduced exactly the way it entered the tube. The effects

The symbols listed below will be used in the following discussion:

- Z_k = sum of all the cathode impedance (R_k and C_k are its components)
- Z_B = the bulk or oxide impedance (R_B and C_B are its components)
- Z_i = the interface impedance (R_i and C_i are its components)
- Z_a = the "peeling" impedance (R_a and C_a are its components)

All present-day instruments measure the resistive components of the cathode impedance. The direct-cur-

rent method measures R_k. The alternating-current method will measure at least R_i and some fraction of R_a and R_B. Pulse methods measure R_i and some fraction of R_a, and R_B. Present instruments were designed solely to measure the interface resistance. We shall, therefore, discuss them and their use in measuring R_i and C_i and then relate them to R_a and R_B. The direct-current method will be omitted because it is time-consuming and, therefore, little used.

Measurements of R_i are designed so as not to change the cathodes in any way. Signal levels must be low, and reduced heater voltages are normally prescribed. It is common practice for most tests of this type to employ signal levels of from 0.1 to 0.3 volt and heater voltages of about 90 per cent of normal operating voltage. It is also recommended[23] that test current be held to approximately five milliamperes per square centimeter of coated area. This current density is chosen to minimize changes in the interface that might be produced by higher cathode-current drains.

When a tube is tested for cathode impedance, it is generally advisable to employ the triode type of test. Most acceptable test methods use this type of connection in preference to the diode connection. In this way only the cathode characteristics are measured. The grid is biased so that it will not draw current, and the plate is grounded with respect to the signal voltage. Otherwise, the effects of a grid or a plate impedance might be included in the impedance measurement.

Sine-Wave Methods

Sine-wave methods generally employ two[24] or three[25, 26] frequencies. A high frequency of 10 megacycles is common to all. At this frequency C_k acts as a short circuit to the signal, and only plate resistance is measured. A low frequency of about five kilocycles is then passed through the tube and C_i or C_a will act as an open circuit to this frequency. Here R_i or R_a or both are evaluated because the entire degeneration is assumed to take place across the resistive component of the impedance. If bulk impedance is present, its resistive component may also be effective at this frequency. A third, intermediate, frequency is often added to evaluate the capacitive component of whatever the system has measured for R_k. These systems inherently assume that a single value of resistance in parallel with a single capacitance will explain the entire behavior of any and all cathode impedances present in the cathode. This assumption is not true. A single two-element impedance can only approximate one impedance at a time. It is a fallacy to lump two or more of them together in the same two-element approximation.

In spite of the obvious discrepancies of the sine-wave methods, the method contributed by Wagner[24] has had some success in evaluating cathode impedance. It uses a bridge type of circuit (Fig. 12) which permits the tube to be measured as a diode or a triode. The input signals are 10 megacycles and 5 kilocycles at 0.1 volt root-mean-square across the tube; the bridge is balanced by the null methods. Wagner's method will give a single value for the cathode resistance but neither C_k nor a time constant can be determined.

Pulse and square-wave systems have been most successful in evaluating cathode impedances. Their biggest advantage is that the effect of the impedance on a pulse signal is normally presented on an oscilloscope. The experimenter can visually estimate the type of impedance by noting the approximate time constant of the decay. Diode and triode tests are both possible but, again, the triode system is preferred. The pulse length used is generally 10 or 20 microseconds with an amplitude of about 0.2 volt peak. The output pulse or square wave should have a rise time no greater than 0.05 microsecond.

Square-Wave Methods

Two square-wave systems currently employed by the tube industry are (1) the Frost[23] method and (2) the Shipley[23] method. In the Frost method illustrated in Fig. 13, combinations of resistance and inductance are inserted in the cathode circuit of the tube under test until two horizontal lines are made to coincide on the oscilloscope screen. From the values of the inserted resistance and inductance, both the resistance and the capacitance of the cathode impedance can be determined. This method is good, but time-consuming.

In the Shipley method, a square is shaped to match the decay produced in another square wave by cathode impedance in the tube under test. The waves are matched by inserting resistance and capacitance in the cathode circuit of a "standard tube." The outputs of the standard and the tube under test are displayed on an oscilloscope via an electronic switch. Both the resistance and the capacitance of the cathode impedance are determined. This system circuit is shown in Fig. 14. For a single cathode impedance, such as R_iC_i or R_aC_a, the approximation is quite good. The approximation has the added advantage that, should there be more than one impedance present in the cathode of the tube under test, the experimenter is readily aware of the condition because the exact decays are visually displayed. The sole objection to this method is the rather dubious entity of a standard tube of universal application which must, of course, be impedance-free and completely stabilized.

Pulse Method

In an attempt to reduce cathode impedance testing to a production basis, and yet maintain a reasonable degree of accuracy, Hannig[27] developed a pulse method which has met with considerable success within RCA. The circuit employs a differential amplifier in which only the decay produced by the cathode impedance is amplified and displayed on an oscilloscope. The system uses a pulse input of 10 microseconds minimum width and a peak amplitude of 0.3 volt. The minimum repetition rate is 5000 pulses per second. The system was originally a triode test but was modified to include a diode test. The diode test uses the same pulse width but an amplitude of about 5 to 8 volts peak. The same differential amplifier serves both tests. The diode test circuit is given in Fig. 15. Values of R_i measured on this system agree closely with those read on the Frost system. R_a and R_B may also be evaluated; evaluation of R_B requires a pulse width of more than 20

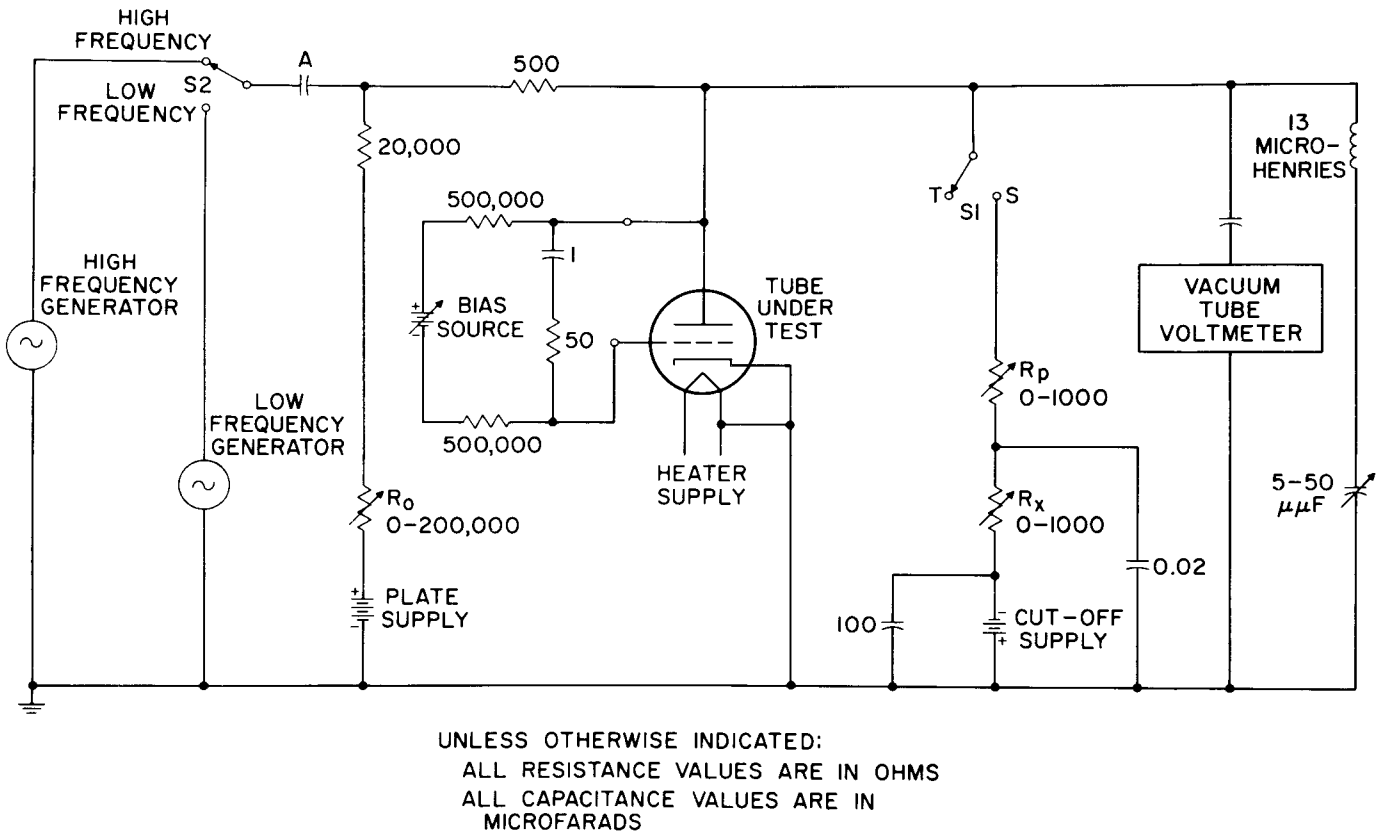


Figure 12. Wagner's Sine-Wave-Method Test Circuit

microseconds. The addition of a variable capacitance in the cathode circuit of the tube under test also gives some measure of the associated capacitance of the impedance. The variable capacitor is calibrated directly in terms of time constants, which are used to separate Z_a, Z_i, and Z_B.

By combining diode and triode testing techniques, it is possible to differentiate between grid and cathode impedances. This method has been applied with excellent results to numerous tube problems. As such, it represents a nondestructive test and allows further testing of the malfunctioning grid or cathode. Drawings of the effects of various tube impedances are given in Fig. 16 (A to E). Table III gives information on each part of Fig. 16.

The diode test might also have been used to yield the data of columns (A), (C), and (D) of Table III. However, had grid deposits been present there is a possibility that the grid impedance effect could have been added to these decays. There is no way to screen this effect out in the diode test. At first glance, columns (B) and (E) appear to be similar. Both are diode tests; both have time constants of less than 10 microseconds. The experimenter can determine which impedance is involved by reference to the fifth and sixth rows. The change in heater voltage from five volts to six volts decreases the magnitude of R_B, whereas this change in heater voltage only serves to increase R_G (resistance of grid deposits) or leave it unchanged. R_i and R_a also show a temperature dependence similar to that of R_B.

Table III

Data	Fig. 16 Reference				
	(A)	(B)	(C)	(D)	(E)
Tube Type	12AU7	6AG7	6AG7	12AT7WA	6AF4A
Cathode Alloy	N109	N81	N34	N109	N18
Test Type	Triode	Diode	Triode	Triode	Diode
Impedance	None	Bulk	Interface	Poor adherence	Grid
Resistance—ohms (6.0 V = E_H)	0	55	180	240	380
Resistance—ohms (5.0 V = E_H)	0	80	230	300	350
Time Constant—microseconds	----	>10	1	0.1	>10

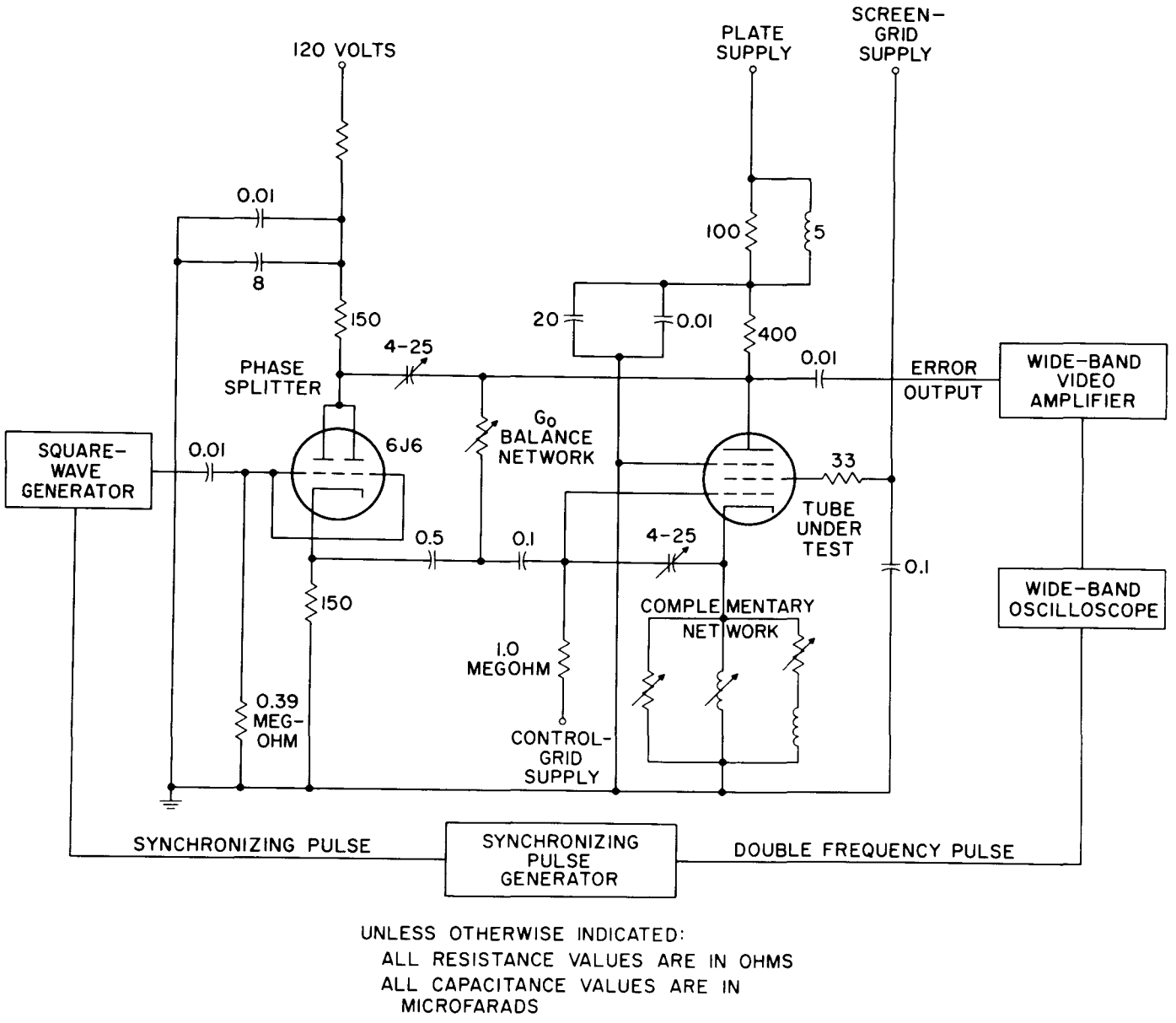


Figure 13. Frost-Method Test Circuit

Test devices which perform the function of evaluating a cathode impedance all operate so that the tube condition is unchanged. To that end, a few "musts" are enumerated to guide in the selection of the best conditions for life-test impedance testing. For cathode impedance testing only on the triode system:

1. Do not test with heater voltage greater than 90 per cent of normal
2. Use low-amplitude test signal
3. Maintain grid bias beyond cutoff; no grid current is permissible
4. Do not allow direct-current cathode current to exceed five milliamperes per square centimeter of cathode area

REFERENCES

1. Wehnelt, A., "Uber den Austritt negativer Ionen aus gluhenden Metallen," Ann. der Phys., Series 14, No. 8, pp. 425-468, 1904
2. Arnold, H. D., "Phenomena in Oxide-Coated Filament Electron Tubes," Phys. Rev., Vol. 16, pp. 70-82, July 1920
3. Eisenstein, A., "Some Properties of the Ba_2SiO_4 Oxide Cathode Interface," Jour. Appl. Phys., Vol. 20, pp. 776-790, August 1949
4. Nergaard, L. S., and R. M. Matheson, "Studies of the Interface Layer in Oxide Cathodes," RCA Review, Vol. 15, pp. 335-361, September 1954
5. Nergaard, L. S., "Studies of the Oxide Cathode," RCA Review, Vol. XIII, pp. 464-545, December 1952
6. Waymouth, J. F., Jr., "Deterioration of Oxide-Coated Cathodes Under Low Duty-Factor Operation," Jour. Appl. Phys., Vol. 22, pp. 80-86, January 1951
7. Eisenstein, A., "A Study of Oxide Cathodes by X-

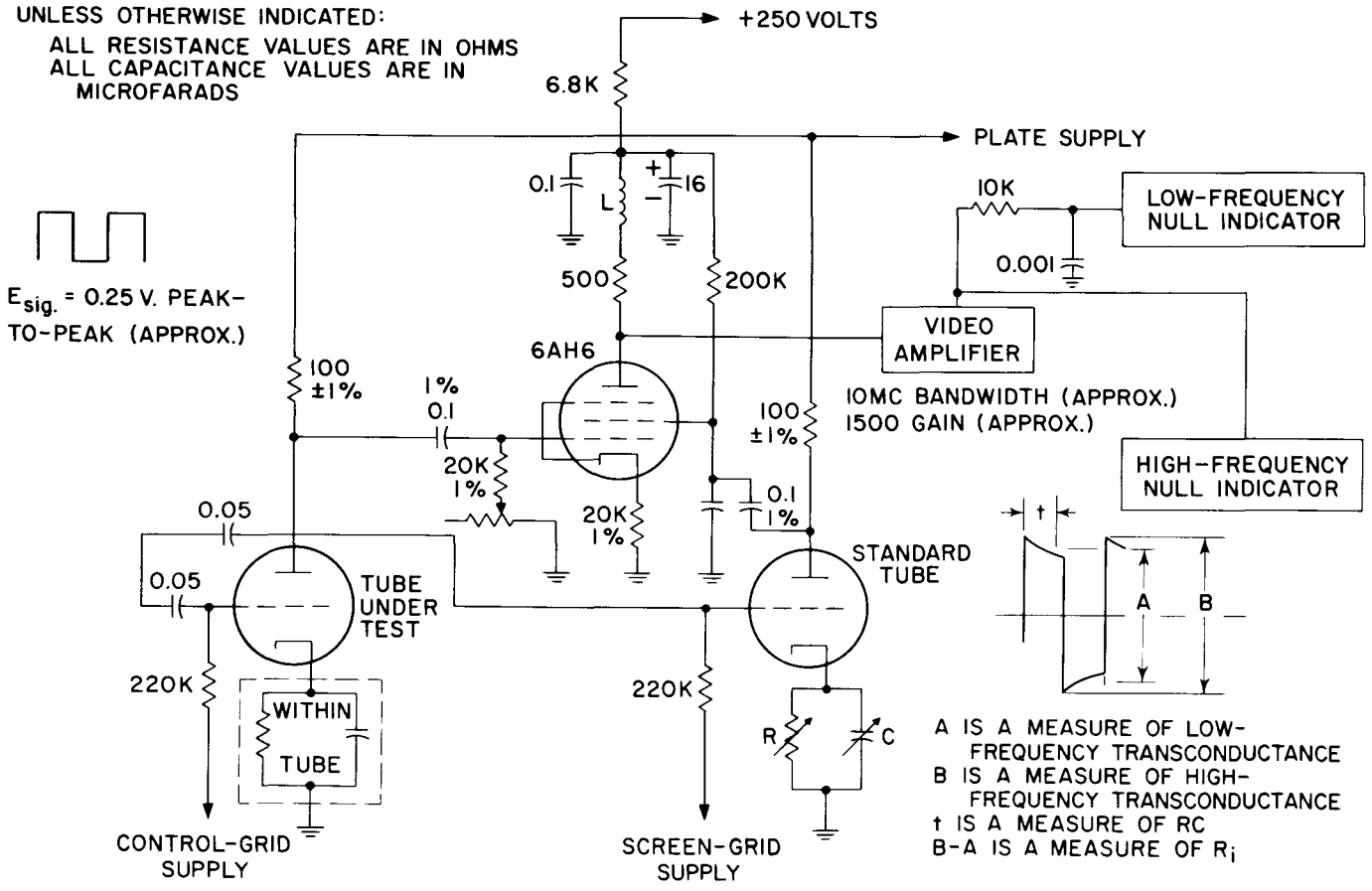


Figure 14. Shipley-Method Test Circuit

ray Diffraction Methods," Part I, Jour. Appl. Phys., Vol. 17, pp. 434-443, June 1946; Part II, Jour. Appl. Phys., Vol. 17, pp. 654-663, August 1946

8. Fineman, A., and A. Eisenstein, "Studies of the Interface of Oxide-Coated Cathodes," Jour. Appl. Phys., Vol. 17, pp. 663-668, August 1946
9. Rooksby, H.P., "Identification by X-rays of Interface Compounds on Oxide Cathodes," Nature, Vol. 159, pp. 609-610, May 1947
10. Eisenstein, A., "The Leaky-Condenser Oxide Cathode Interface," Jour. Appl. Phys., Vol. 22, pp. 138-148, February 1951
11. Bounds, A. M., and T. H. Briggs, "Nickel Alloys for Oxide-Coated Cathodes," Proc. IRE, Vol. 39, pp. 778-779, July 1951
12. "Sublimation Characteristics of Metallic Materials by Electrical Resistance," A.S.T.M. Spec. F-278-55T
13. Haase, A. P., and J. W. Denison, "Some Sublimation Characteristics of Cathode Nickels," Rpt. on the Twelfth Ann. Conf. on Phys. Electronics, M.I.T., pp. 51-57, March 1952
14. Rittner, E.S., "A Theoretical Study of the Chemistry of the Oxide Cathode," Philips Research Rpts., 8, pp. 184-238, June 1953
15. Nergaard, L. S., and R. M. Matheson, "A Bridge for the Measurement of Cathode Impedance," RCA Review, Vol. 15, pp. 485-505, December 1954
16. Code T1
17. Raudorf, W., "Change of Mutual Conductance with Frequency," Wireless Engineer, Vol. XXVI, No. 313, pp. 331-337, October 1949
18. Sproul, R. L., "An Investigation of Short-Time Thermionic Emission From Oxide Coated Cathodes," Phys. Rev., Vol. 67, pp. 168-178, March 1 and 15, 1945
19. Coomes, E. A., "The Pulsed Properties of Oxide Cathodes," Jour. Appl. Phys., Vol. 17, pp. 647-654, August 1946
20. Seymour, J., "Computer Valves and Cathode Interface Impedance," Jour. of Electronics and Control, Vol. 3, No. 1, pp. 107 ff, July 1957
21. Metson, G.H., "On the Electrical Life of an Oxide-Cathode Receiving Tube," Advances in Electronics and Electron Physics, Vol. VIII, pp. 403-446, Academic Press
22. Dalman, G. C., "Effects of Cathode and Anode Resistance on the Retarding Potential Characteristics of Diodes," Rpt. on the Thirteenth Ann. Conf. on Phys. Electronics, M.I.T., pp. 106-112, 1953
23. "Interface Impedance Characteristics of Vacuum Tube Cathodes," A.S.T.M. Spec. F-300-55T, June 1955
24. Wagner, H. M., "Cathode Interface Resistance Test for Electron Tubes," Rpt. on the Fourteenth Ann. Conf. on Phys. Electronics, M.I.T., pp. 77-81, 1954

25. Englesfield, C. C., and P. E. Douglas, "A Method of Measuring the Interface Resistance and Capacitance of Oxide Cathodes," Br. Jour. Appl. Phys., Vol. 2, pp. 318-320, July 1951

26. L. R. McNarry Method, "Progress Report ERA-254," National Research Council of Canada, Radio and Electrical Engineering Division, July 1953

27. Code T2

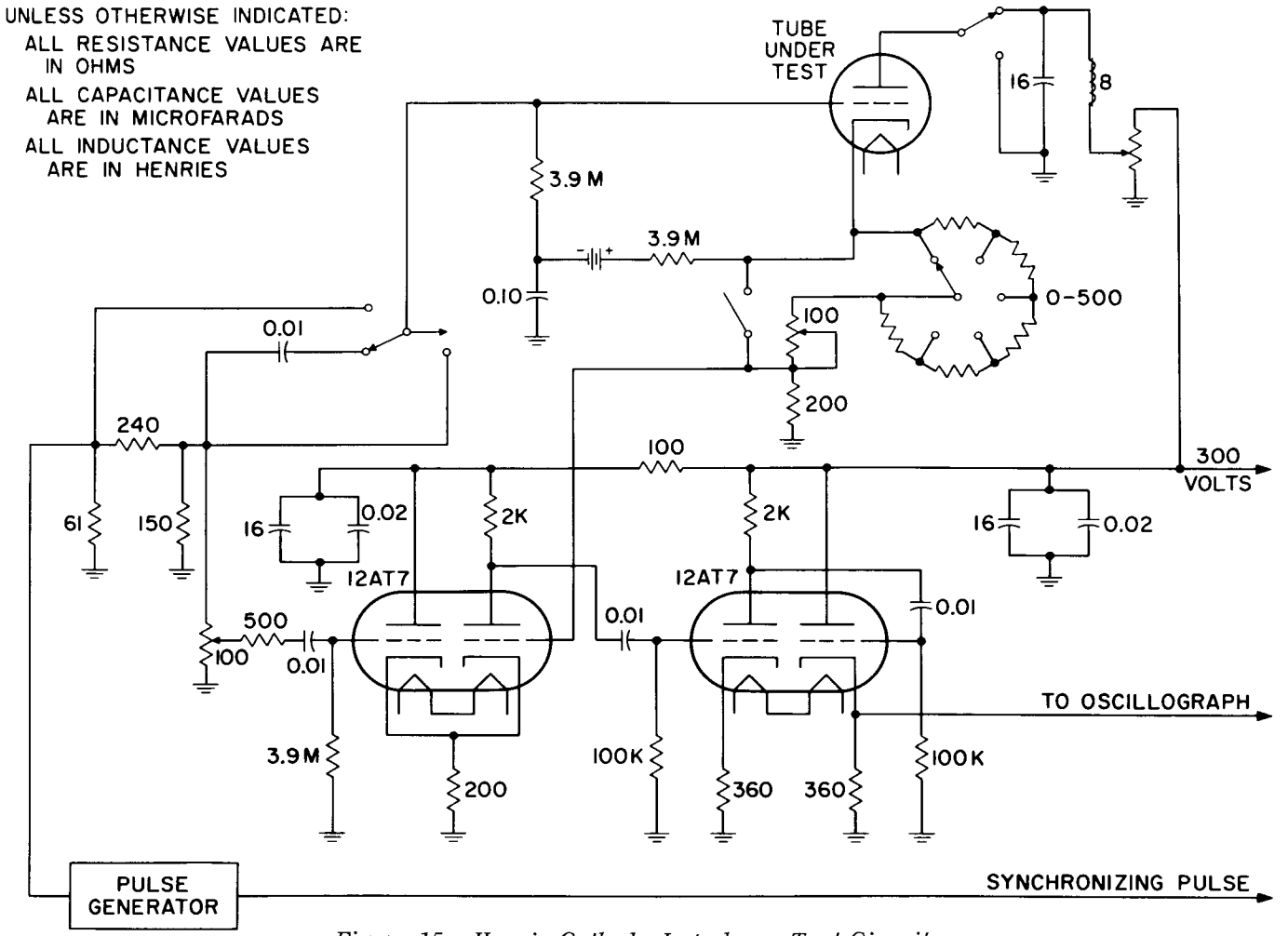


Figure 15. Hannig Cathode-Impedance Test Circuit

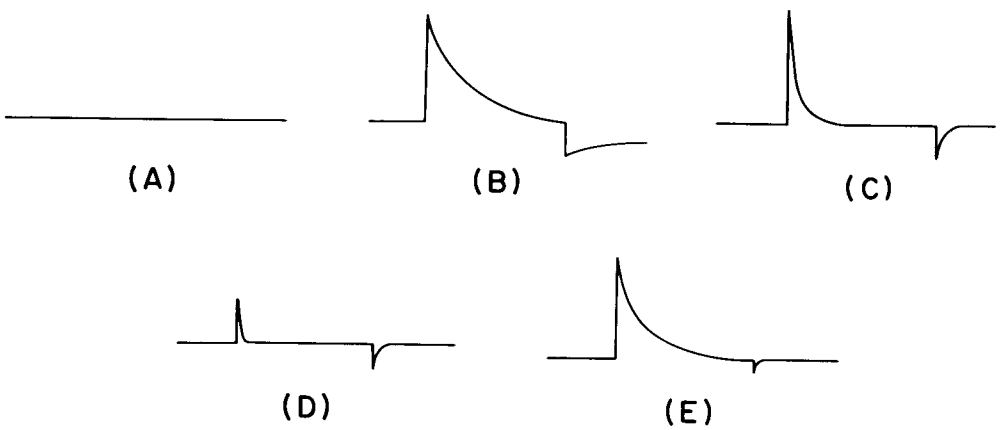


Figure 16. Oscillograph Traces of Square-Wave Distortion Caused by Cathode Impedance: (A) no impedance; (B) bulk impedance; (C) interface impedance; (D) peeling-cathode impedance; (E) grid impedance

Tungsten, Thoriated-Tungsten, and Thoria Emitters

W. E. Harbaugh

Lancaster

PURE-TUNGSTEN EMITTERS

Tungsten has been used for many years as the source of electrons in high-power tubes. Until relatively recently, all tubes operating at plate voltages above 5000 volts used pure-tungsten emitters because of their resistance to high-energy gas ions, which attack oxide-coated and thoriated-tungsten cathodes.

Although tungsten has the relatively high work function of 4.56 volts, its high melting point (3655 K) and great mechanical strength make it possible to operate cathodes at temperatures high enough to achieve practical emission levels.

Preparation of Tungsten Filaments

Ductile tungsten was first produced by Coolidge in 1909. Because of its high melting point, powdered-metallurgy techniques were employed and are still in use today. Tungsten powders are compressed into bars under pressures of about 30 tons per square inch at room temperature, and the bars are then presintered in a furnace at about 1250 C. These bars, which are porous and weak but sufficiently strong to be handled, are then sintered by passage of several thousand amperes through them to heat the tungsten almost to its melting point. During this process the metal powder sinters and the volume shrinks about 17 per cent. The tungsten bar is then very strong, but quite brittle. In the next step, it is hot-swaged down to a small rod size by rotating hammers. Swaging produces an elongated crystal structure which makes the metal ductile.

The tungsten is then ready for drawing or rolling to produce the wire sizes used in filamentary cathodes, The tensile strength of wire produced in this manner may reach 500,000 pounds per square inch; such wire is reasonably ductile. Required "bends" can be made cold, but severe working is best accomplished at temperatures between 800 and 1000 C. The long fibrous structure of pure (undoped) tungsten remains until the wire is heated above 1200 C. At this temperature the metal begins to recrystallize into a fine-grain structure and suffers a loss of ductility. At the high temperatures (2400 K) required for electron emission, the grains continue to grow and eventually form large crystals which cause the filaments to become extremely fragile. In some instances the crystals may extend completely across the diameter of a filament, and the resulting offsetting may cause the filament to break mechanically long before failure due to normal evaporation occurs. Efforts to produce single-crystal wire and "doped" wire to prevent grain growth and offsetting have been moderately successful. Thoria was used as an early addition to pure tungsten to inhibit grain growth; the fact that it also greatly enhanced the electron-emission properties of tungsten proved to be an unexpected extra dividend.

The grain size of tungsten is of considerable importance when the wire is severely worked. When round wire is rolled into strands having odd-shaped cross-sectional configurations, wire having 1000 grains per square millimeter, since it works more easily than coarser-grained material, is far better for achieving uniform dimensions. However, fine-grain tungsten has low creep strength so that, for maximum life, tungsten having large grain structure (less than 500 grains per square millimeter) is desirable. Grains should have a length-to-width ratio of at least 5 to 1 to achieve high strength. Because the grain size is controlled by the tungsten supplier, close cooperation between the maker and the user of tungsten is very desirable for achieving a proper balance between workability and strength.

Properties of Tungsten Emitters

Electrons are emitted from tungsten when sufficient thermal energy is induced to cause the electron to cross the surface-boundary region. The properties of tungsten, its emission characteristics, and the emission properites of various materials have been thoroughly studied and are well covered in the literature[1-20].

Emission from pure metals such as tungsten follows the Richardson-Dushman equation as closely as can be determined experimentally. This relationship is:

$$I_S = AT^2 \epsilon^{-b_O/T} \qquad (1)$$

where I_S = Emission current — amperes
A = Constant for the emitting material — amperes/square centimeter/deg K
T = Absolute temperature — deg K
b_O = Constant corresponding to the temperature required for all of the free electrons in the material to be emitted — deg K

Table I shows the values of A, work function ϕ, and b_O for several pure-metal emitters; thoriated tungsten is included for comparison.

TABLE I

Emission Factors for Tungsten and Some Other Metals

Emitter	A amp/cm^2/deg^2	Work Function ϕ Volts	b_o degrees K
W	60-100	4.54	52,400
Ta	60	4.1	47,200
Mo	60	4.15	51,300
Th on W	3.0	2.63	30,500

The thermionic properties of tungsten are well known; Table II shows some of these properties which are important to tube designers.[21] In this table, there are certain important properties to be noted. First, it is apparent that there is a large increase in resistance with temperature. This increase must be accommodated by heating the tungsten filament slowly to prevent burnouts caused by sudden surges of current. (This precaution also applies to thoriated tungsten, although to a lesser degree.) Second, the power radiated is so high at practical emission levels that this type of filament is basically inefficient. At a given emission level, pure tungsten requires about six times as much heating power as thoriated tungsten, and about 30 times as much as oxide-coated cathodes.

Table II also shows that the emission level is a steeply rising function of temperature; however, the evaporation-rate rise, shown in the last column, is even steeper. The evaporation rate becomes a critical factor as the operating temperature increases.

Filament life is assumed to end when 10 per cent of the metal has evaporated, because hot spots and burnouts usually occur at this point. Tubes having filaments of less than 0.010-inch diameter are generally operated at approximately 2400 K. Tubes having filaments of larger diameter can be successfully operated up to 2800 K because the percentage loss due to evaporation becomes smaller with increasing bulk of the emitters.

Oxygen increases the work function of tungsten to 9.2. Other electronegative elements also reduce the emission capabilities of tungsten. For this reason, tungsten emitters must be clean to achieve design-emission levels. A tungsten filament is usually cleaned by heating it to somewhat above its operating temperature for approximately 30 minutes to degas the material throughout. A final flash at 3000 K for approximately one minute removes all the surface oxide and provides a clean emitting surface. This treatment is usually given to the filaments during exhaust and processing of the assembled tubes.

Present Uses

Recent work[22, 23] has shown that the more efficient thoriated-tungsten filament can be used at high voltages and in high-power tubes; new tubes, therefore, are no longer being designed with pure-tungsten emitters. However, several tubes using pure-tungsten emitters, such as the 207, 891, 892, 880, 9C21, and 9C22 are still in production. Pure-tungsten emitters are also used in tubes, such as continuously pumped Resnatrons for particle accelerators, where the degree of vacuum is too poor for thoriated-tungsten operation.

THORIATED-TUNGSTEN EMITTERS

Langmuir[24] discovered that thoria (thorium oxide), previously mixed with tungsten to reduce grain growth in tungsten filaments, could greatly increase thermionic emission under proper conditions. It was found that metallic thorium, produced by reduction at high temperatures, diffused to the surface of the tungsten to produce a layer one-molecule thick. It was further found that the electropositive dipole effect of this monomolecular layer reduced the work function of the tungsten to 2.63, which is less than the value of 3.35 for pure thorium.

Preparation

Thorium oxide is added to the tungsten powder in proportions between 0.7 and 2 per cent. The finished thor-

TABLE II

Thermionic Properties of Tungsten at Various Temperatures

Temperature deg K	Resistivity microhms/cm	Total Radiation Intensity watts/ cm^2	Electron Emission amp/cm^2	Rate of Vaporization grams/cm^2/sec
300	5.64	0.0015	--	--
600	13.54	0.048	--	--
900	22.58	0.379	--	--
1200	37.02	1.691	--	--
1400	38.52	3.82	5.75×10^{-9}	--
1600	45.22	7.77	8.05×10^{-9}	3.7×10^{-20}
1800	52.08	14.22	3.92×10^{-5}	6.22×10^{-17}
2000	59.10	23.72	8.92×10^{-4}	2.32×10^{-14}
2200	66.25	37.18	1.14×10^{-2}	2.90×10^{-12}
2400	73.55	55.8	1.02×10^{-1}	1.58×10^{-10}
2600	81.0	80.8	6.48×10^{-1}	4.64×10^{-9}
2800	88.5	112.9	3.21	8.28×10^{-8}
3000	96.2	153.9	--	9.92×10^{-7}

iated filament must be flashed to a high temperature (2800 K) to reduce some of the thoria to thorium metal. At this temperature thorium diffuses to the surface and rapidly evaporates so that the emission at high temperature is essentially that of pure tungsten. If the temperature is then reduced to between 2000 and 2200 K, the rate of diffusion is still quite high, but the rate of evaporation is reduced to the point where a monomolecular layer can form on the surface. This treatment is known as activation and usually is accomplished in 15 to 30 minutes during tube processing. The filament temperature may then be reduced to the operating range of 1900 to 2050 K, where generous electron emission occurs and the rate of thorium evaporation is substantially equal to the rate of diffusion to the surface.

Thorium apparently diffuses to the surface along grain boundaries and spreads over the surface by migration. The surface coverage depends on the orientation of the tungsten crystals; certain faces show preferential adsorption. It is also believed that finer-grain material is beneficial because many more paths are available for outward diffusion of thorium and, therefore, coverage can be accomplished more quickly.

Reasons for Carburization

It has long been the practice to form a layer of tungsten carbide on the surface of thoriated-tungsten filaments. Tungsten carbide greatly reduces the rate of evaporation of thorium from the surface[25]; at 2200 K, for instance, the rate of evaporation is reduced 83 per cent. In addition, the fine grain structure produced by the preferred type of carbide assists rapid diffusion of thorium to the surface. Higher-temperature operation and greater thorium mobility make the carburized filament much more rugged in the presence of gas ions.

This rugged carbide layer, together with the better vacuums and cleaner parts of modern tubes, and the increased protection offered by "crowbar" circuits[26], have made it possible to use thoriated-tungsten filaments in tubes having plate voltages as high as 40,000 volts.

According to Ayer, the life of a carburized filament is essentially a function of carbide thickness, as shown in Fig. 1. More recent information indicates Ayer's curves to be conservative, although they are still highly useful in tube design.

The Carburization Process

The carbide layer is formed by heating the filament in a carboniferous atmosphere so that carbon reacts with the tungsten surface to form a shell of carbide around the filament. Past practice has been simply to carburize until a certain decrease in resistance of the filament showed that a certain percentage of the filament cross section was carburized. A 20-per-cent reduction in resistance was the average figure used.

It has been found most practical to carburize the filaments before they are inserted into tubes by heating them in a bell jar containing dry hydrogen and a hydrocarbon such as benzene or toluene. When the hydrocarbon vapor strikes the hot filament it dissociates, and at high temperature the carbon reacts with the tungsten to form tungsten carbide. A typical apparatus for performing this operation is shown in Fig. 2.

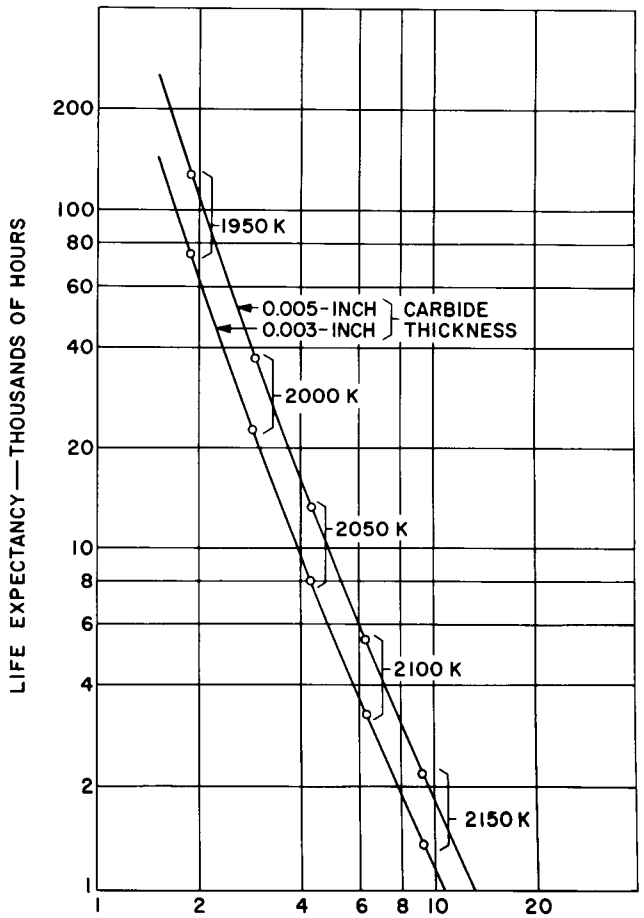


Figure 1. Life Expectancy and Emission for Thoriated-Tungsten Filaments for two Thicknesses of Carbide

Many carbides are possible in such a setup and, therefore, careful control of all parameters must be maintained to achieve the desired results. In 1947, Horsting[27] first showed that the carbon potential of the hydrogen atmosphere was important, not only for maintaining a uniform depth of carbide, but also for controlling the type of carbide. The various types of carbide possible for different carbon potentials when other conditions remain constant are shown in Fig. 3. There are many forms of tungsten carbide, all grouped around two chemical compounds, WC at 6.12 per cent carbon and W_2C at 3.16 per cent carbon, as shown in Fig. 3. The diagrams at the left of the figure illustrate a pie-shaped, cross-sectional area of a round wire. The figures at the right show the concentrations of carbon in tungsten in respective layers and their relative depth of penetration.

It can be seen that the stoichiometric WC produces a massive, unbroken carbide; W_2C is also a massive carbide but breaks up on radial lines. As the carbon concentration diminishes, a laminar carbide is formed in the region between 3.16 per cent C (W_2C) and 2.45 per

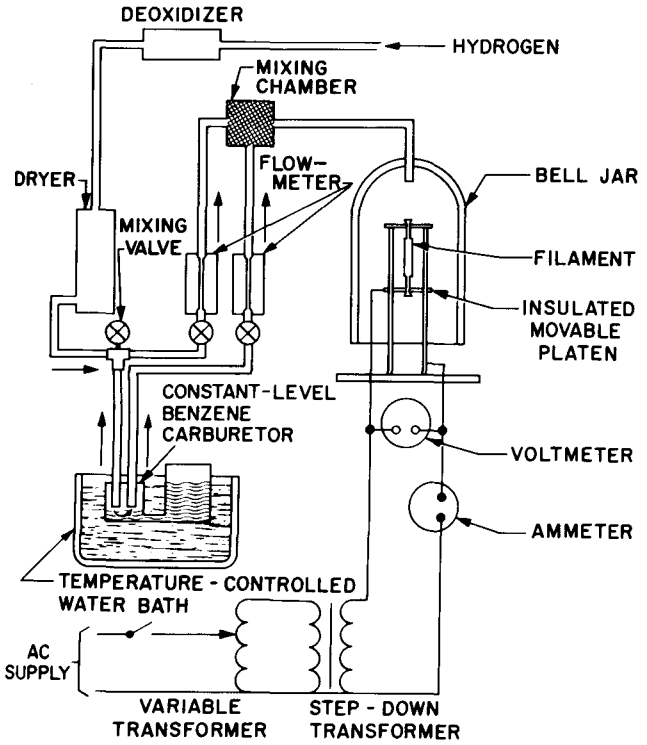


Figure 2. Typical Equipment for Carburizing Thoriated-Tungsten Filaments

cent C. This type of carbide is caused by a narrow region of solubility of tungsten in W_2C which precipitates out on cooling to form a laminated structure. Further reductions in carbon potential reduce the penetration but still maintain the laminar structure.

Further work[28] has indicated that the finely laminated structure, as shown in F of Fig. 3, is the desirable type for full, stable emission in high-power tubes. It is believed that the broken-up carbide structure allows the thorium to diffuse to the surface more readily and it is therefore available at all times for electron emission.

The typical cross section of a filament strand used in an RCA Super-Power tube in Fig. 4 shows the preferred type of laminar carbide. An example of the undesirable, massive type of carbide is also shown.

In most coiled filaments, the carbide is produced in a bell jar similar to that shown in Fig. 2. The filament is carburized to a specified per-cent change in resistance and is then mounted in the tube. The tube is then exhausted. During exhaust the filament is flashed and activated by the procedures mentioned previously. It has been found, in studies of this action, that a massive-type carbide is formed during carburization and is later reduced to the laminated phase by the flashing and activation processes. The carbon content of the carbide is reduced both by evaporation of carbon and by further penetration at these high temperatures. This method produces satisfactory filaments, but requires very close control of all parameters.

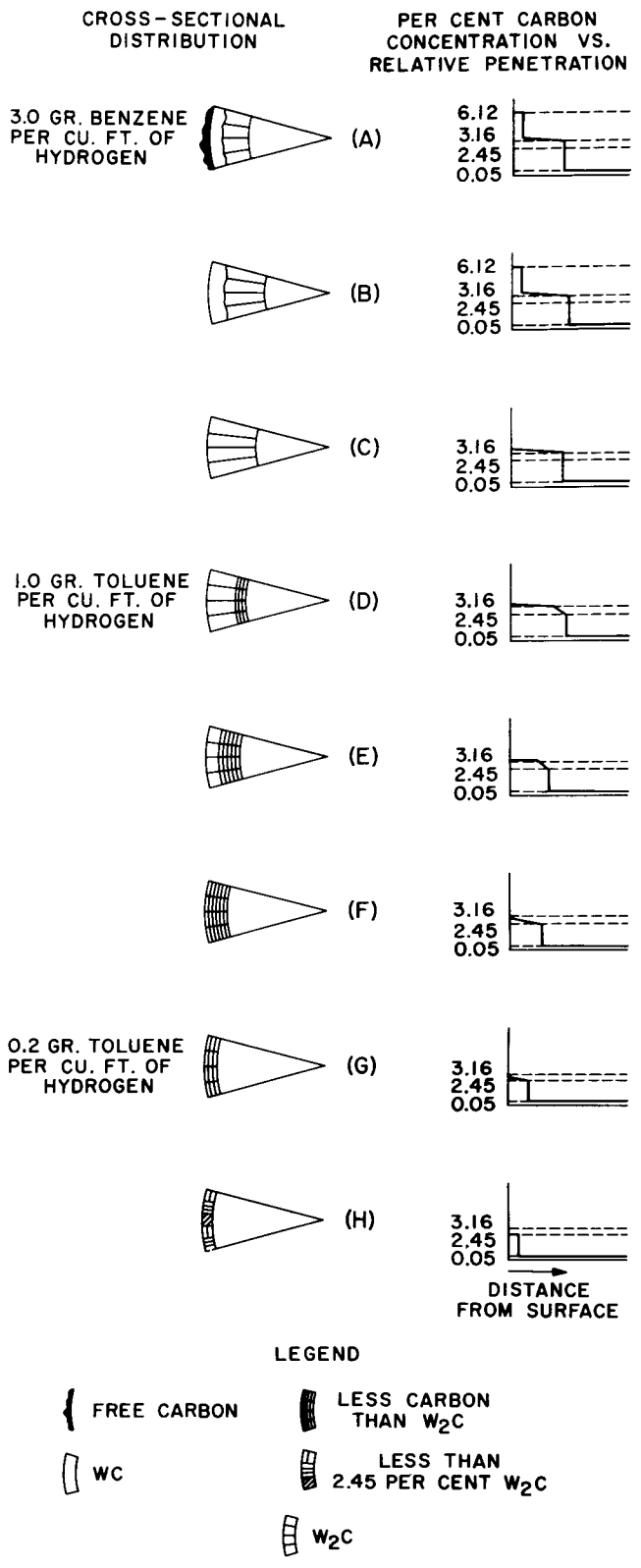
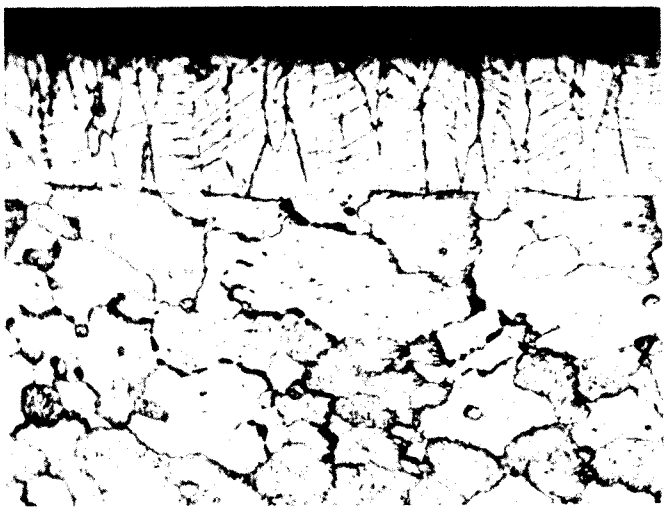


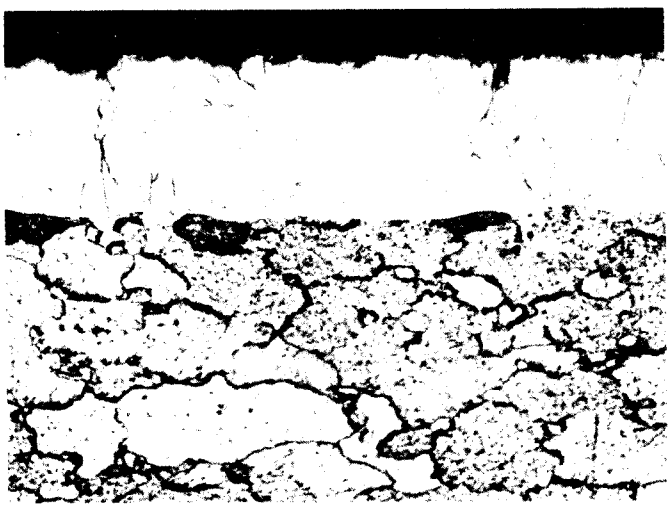
Figure 3. Distribution of Tungsten Carbides on Carburized Thoriated-Tungsten Wire

End-Loss Correction

In the design of straight-strand filaments for Super-Power tubes it was found that any system involving high-



(A)



(B)

Figure 4. Cross Sections of Carburized Thoriated-Tungsten Filaments: (A) preferred type of laminar carbide (272X); (B) undesirable massive type of carbide (272X)

temperature flashing of the filaments during exhaust would not be feasible. This limitation is the price paid for a filament feature which is very important in the design of high-power, high-frequency tubes.

All directly heated cathodes (filaments) must be so mounted that heating current can be passed through them while they are held in proper relationship to other tube elements. In any structure, the heat lost to the mounting mechanism is considerable. To offset this loss, a method known as "end-loss correction" is used on all Super-Power tubes. This correction consists simply of reducing the cross section of the filament near the ends where it is mounted. This arrangement creates sections of higher resistance in which more power is dissipated (because the current through the strand is constant). By proper design, this additional power can be made just sufficient to offset the heat lost to the mountings.

The difference between a corrected and an uncorrected filament is shown diagrammatically in Fig. 5. For a given length, much more filament area is maintained at a useful emitting temperature, or, for a given emitting surface, the tube can be made much shorter, thus greatly enhancing higher-frequency operation.

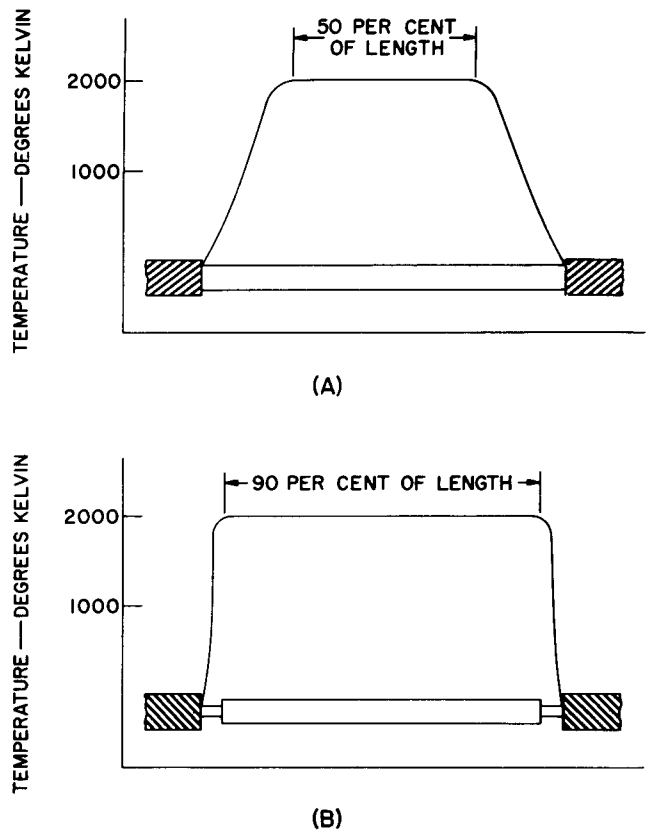


Figure 5. Temperature-Distribution Curves: (A) filament without end-loss correction; (B) filament with end-loss correction

End-loss correction, as described, is a useful tool in the design of high-power tubes. However, it must be pointed out that such a correction can be made for only one temperature. It is normal to correct for the operating temperature. At higher temperatures the thinned cross sections overheat, as shown in Fig. 6. Obviously, such a filament, properly corrected for operating temperature, cannot be flashed at high temperatures during exhaust, without melting the ends.

Preferred Carburization

Experiments have shown that a thoriated-tungsten filament can be carburized directly to the laminated phase in a hydrogen-hydrocarbon atmosphere, and that

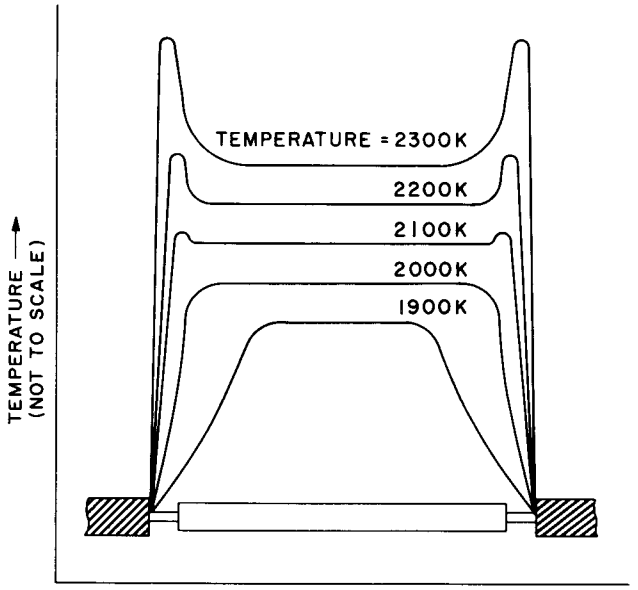


Figure 6. Effect of Various Center Temperatures on a Filament Having End-Loss Correction for an Operating Temperature of 2000 K.

a filament so produced will have full, stable emission at operating temperature without further high-temperature activation. In this method, jets of pure hydrogen are directed at the reduced-area end sections to keep them from burning out or carburizing during the process. The carburizing schedule used on this type of filament is shown in Fig. 7.

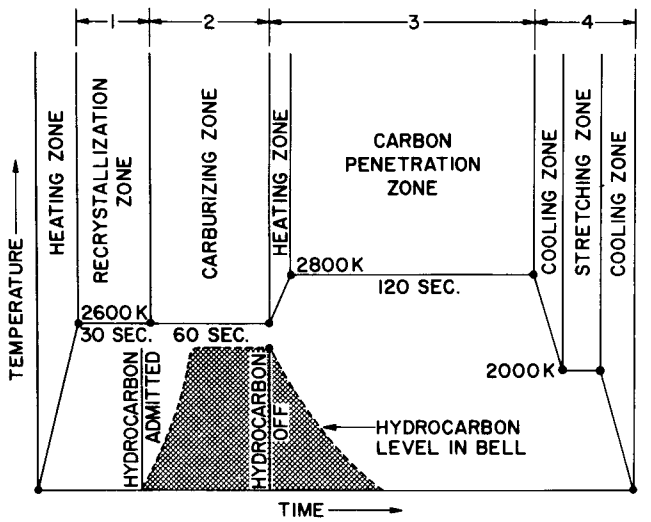


Figure 7. Typical Cycle for Carburizing Filaments

This carburizing process involves four steps: (1) the filament is flashed at a high temperature in pure hydrogen to cause recrystallization and to reduce the thoria to thorium metal; (2) a closely controlled hydrogen-benzene mixture is admitted while the filament is still hot (this step forms a thin rind of massive-type carbide); (3) the flow of benzene is turned off so that no additional carbon is available, and the temperature is increased so that the carbide already formed penetrates deeper and produces the carbon-starved conditions that yield the proper laminated carbide; (4) the filament is stretched approximately one per cent beyond the elastic limit at 2000 K. The stretching step, possible only on straight wire, not only produces a straight filament, important in electron-optical devices, but also appears to enhance electron emission, possibly by pulling apart the grains slightly and allowing more freedom for thorium migration.

Filaments produced by this schedule are presently used in tubes which have been operated more than 20,000 hours without loss of emission. Checks of the carbide have indicated that half of the carbide still remains for future life.

Some Precautions for Carburization

A word of caution must be given regarding the setting up of carburizing schedules. The filament being carburized must be heated in a clean atmosphere, after which the carbon can be added to enter into the reaction. When the filament is heated in a hydrogen-hydrocarbon atmosphere, the hydrocarbon dissociates long before any reaction between the tungsten and carbon occurs. A carbon deposit then forms on the surface which greatly increases the thermal emissivity so that for a given power input the temperature is quite low. In addition, it acts as a shield to prevent additional carbon from reacting with the tungsten. For these reasons, the filament is flashed in pure hydrogen first and held at temperature while the hydrocarbon is added.

The foregoing procedure is recommended; however, particular attention must be paid to all operating conditions because the process is very critical. The amount of hydrocarbon vapor must be closely controlled, all times and temperatures in the process must be held accurately, and the hydrogen must be pure and extremely dry to prevent any reaction with the monolayer of thorium or with the carbide.

Quality Control

The quality of filaments produced by the preceding method is controlled in several ways. The decrease in current (or increase in voltage) indicates the change in resistance during carburization and is one of the conditions kept constant. With many production tubes, thyratrons are used to shut off the heat when the proper resistance is reached. However, resistance is only one factor and can change with surface condition, as pointed out by Horsting[27]. A second check can be made by weighing the filaments before and after carburizing to determine whether they have reacted with the correct amount of carbon. Two types of destructive tests are used, on a sampling basis, to assure the correct carbide structure and thickness. The sample filaments are sectioned and viewed through the metallurgical microscope to study the carbide structure

and measure the layer thickness. In addition, sections of the filaments are placed in a spectrograph and their surfaces explored for thorium. A properly activated filament shows traces of thorium when the surface is arced in a spectrograph.

Electron-emission tests are also made in either finished tubes or in test bottles which have been properly baked and processed, and total space current is plotted against voltage. Fig. 8 shows a plot of voltage and current on a typical RCA Super-Power tube. In Fig. 8, the point where the current begins to break away from the 3/2-slope of the voltage indicates the maximum space-charge-limited emission available.

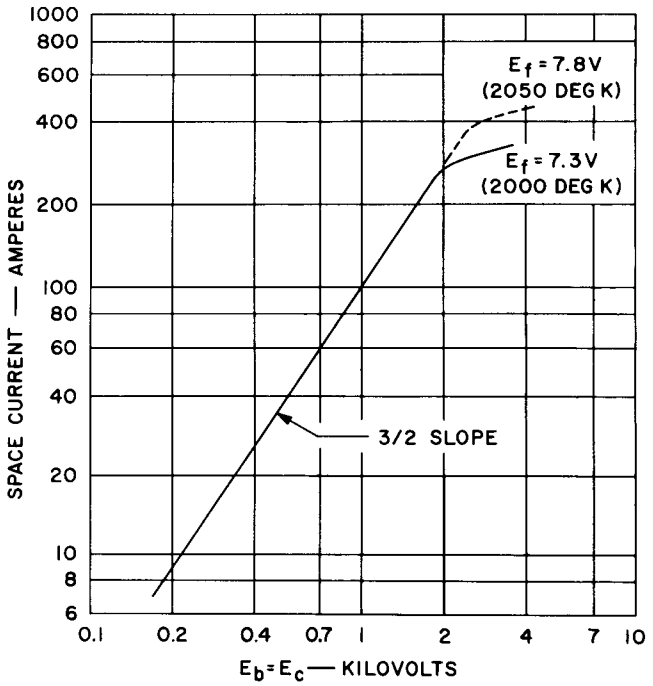


Figure 8. Typical Emission Test of RCA Super-Power Tube; Break-Away Points Indicate Where Cathode Becomes Temperature-Limited

In Fig. 9, a plot of emission versus temperature is shown together with data on the radiation and emission efficiency. Most large power tubes are designed to operate at a filament temperature of 2000K which yields about 3 amperes of emission per square centimeter of emitting area.

Fig. 10 shows curves for use in converting brightness temperature, as read on an optical pyrometer, to true temperature. These curves, which are based on the emissivity of pure tungsten and carburized tungsten, are taken from published literature[29] and hold true for most cases. The curve shown in Fig. 10B shows recent measurements of the spectral emissivity of cathodes used in certain Super-Power tubes. It can be seen that this curve differs slightly from that shown in Fig. 10A. More accurate results can be obtained by measuring the spectral emissivity for each type of cathode to account for surface differences.

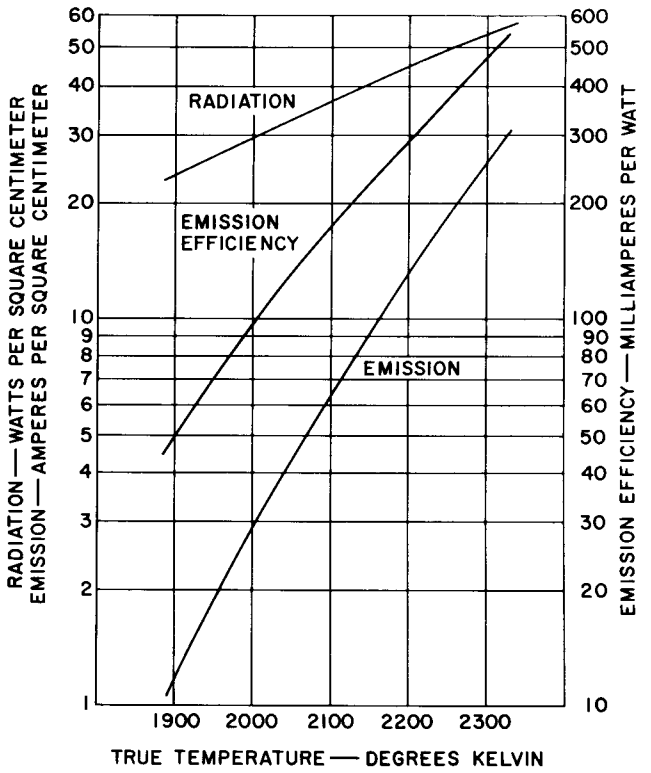


Figure 9. Variation of Emission, Emission Efficiency, and Radiation of Thoriated Tungsten as a Function of Temperature

THORIA EMITTERS

Because thoriated tungsten had proved so attractive as an emitter of electrons, a substantial program was conducted from 1940 to 1950 to achieve electron emission directly from thorium oxide. Thoria has three inherent advantages over thoriated tungsten. First, more shapes of tube elements are possible because of the ease of working the base metals to which thoria is applied. Second, higher emission densities seem possible with thoria, although this increase may be due to the increased surface area. Third, thoria cathodes are easy to activate and are less susceptible to poisoning. They appear to be extremely rugged and continue to emit electrons after treatments which completely kill thoriated tungsten or barium-strontium oxide.

Thorium-oxide powder (ThO_2) is usually coated on a refractory base metal such as tungsten (W), tantalum (Ta), or molybdenum (Mo) by painting, spraying, or cataphoresis. The base metal reduces some of the thorium oxide to thorium, which then diffuses to the surface. Emission occurs from a monolayer in a manner similar to that of thoriated tungsten.

Forgue[30] experimented with many combinations of metal powders sintered to a base metal and coated with thoria (THO_2 on W on Ta, ThO_2 on Mo on Ta, ThO_2 on W on Mo, etc.). A tantalum base metal was preferred because it could be shaped easily; the recommended cathode at that time was ThO_2 on Mo on Ta. Fig. 11 compares Forgue's emission data for two experimental thoria combinations with that for pure tungsten, pure tantalum, and thoriated tungsten.

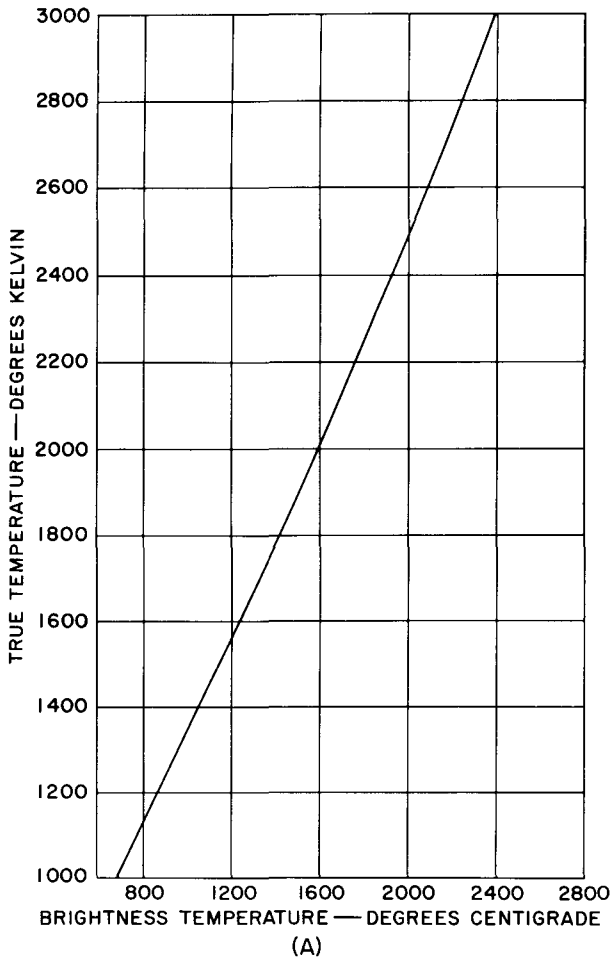


Figure 10. Curves for Conversion of Brightness Temperature to True Temperature: (A) pure tungsten (for carburized tungsten, curve is about 10 degrees below that shown and parallel to it); (B) carburized thoriated tungsten (as measured on a super power tube)

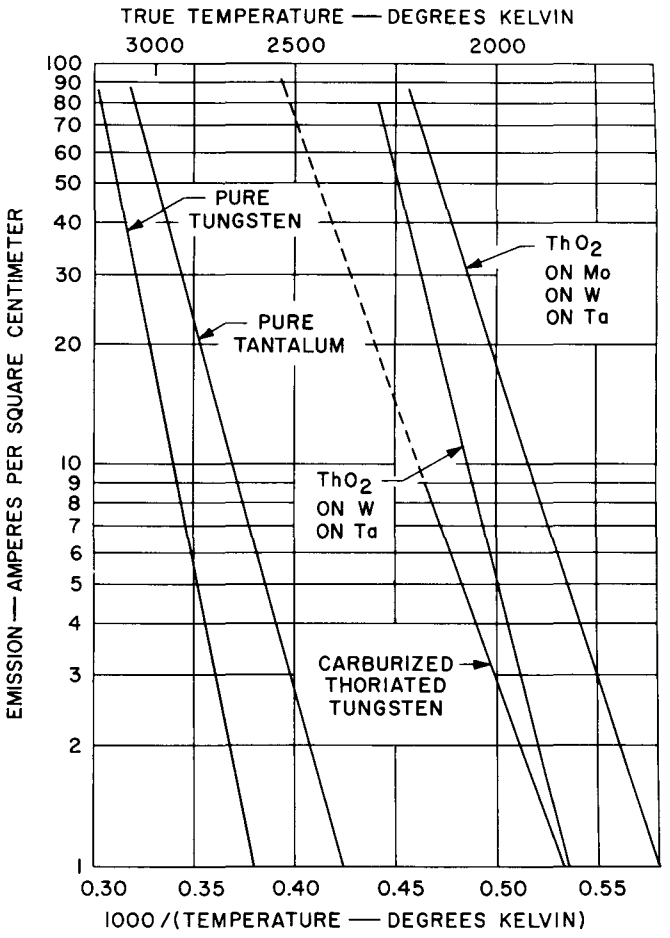


Figure 11. Comparison of Thoria Emission-Temperature Data (from Forgue) with Data for Standard Cathodes

Thoria cathodes have not proved to be the panacea once expected. The principal difficulty has been in keeping the thorium oxide on the base metal. The oxide is not only difficult to keep on mechanically, but is also easily ripped loose under severe ion bombardment.

The solution to these mechanical problems may be in the area of "cermet" or "dispenser" type cathodes. Tubes have been made with directly heated thoria cathodes pressed into cylindrical form and made conductive by addition of tungsten or molybdenum powder to the thoria before pressing [31].

Thoriated tungsten is still the more reliable emitter although, with further effort, thorium oxide may well displace it in the future. At present, only one commercial RCA tube employs the thoria cathode: the 8D21 vhf television tetrode. Life and reliability now appear to be excellent on this tube type, although early production was beset by many cathode difficulties. Thoria cathodes have many attractive attributes. It is possible that matrix technique, where the emissive material is held within a porous metal layer, may well improve the mechanical structure to the point where thoria will become a reliable cathode material.

SUMMARY

Pure-tungsten cathodes are no longer being designed into high power electron tubes. The penalty paid for high cathode heating power is no longer justified since thoriated-tungsten emitters have been improved to the point where reliable operation in even the highest power tubes is now possible. Thoriated tungsten must be properly made and the processing carefully controlled to achieve the proper grain structure and carbide layer for full uniform electron emission under extreme voltage conditions. Although thoriated tungsten is used almost universally in all high-power electron tubes, thoria-coated cathodes have excellent possibilities and may, in the future, displace thoriated tungsten as the preferred high-power emitting material. Many problems are yet to be solved to advance the state of the art of this material; however, there is no fundamental reason why it should not be the cathode material of the future.

REFERENCES

1. Richardson, O. W., Emission of Electricity from Hot Bodies, Longmans, Green & Co., New York, 1916, rev. 1921
2. Sommerfeld, A., "The Electron Theory of Metals Based on Fermi Statistics," I. Zeits F. Physik, 47, 1, 1928
3. Dushman, S., "Electron Emission from Metals as a Function of Temperature," Phys. Rev., 21, 623, 1923
4. Bridgman, P. W., "A Critical Thermodynamic Discussion of the Volta, Thermo-Electric and Thermionic Effects," Phys. Rev., 14, 306, 1919; "The Universal Constant of Thermionic Emission," Phys. Rev., 27, 173, 1926
5. Compton, K. T., and I. Langmuir, "Electrical Discharges in Gases," Rev. Modern Phys., 2, 137, 1930
6. Davisson, C., and L. H. Germer, "The Thermionic Work Function of Tungsten," Phys. Rev, 20, 300, 1922
7. Waterman, A. T., "The Variation of Thermionic Emission with Temperature and the Concentration of Free Electrons within Conductors," Phys. Rev., 24, 366, 1924
8. Dushman, S., H. N. Rowe, J. Ewald, and C. A. Kidner, "Electron Emission from Tungsten, Molybdenum and Tantalum," Phys. Rev., 25, 338, 1925
9. Schottky, W., "On the Influence of Structure Effect, Especially the Thomson's Image Force, on Electron Emission from Metals," Physik Zeit, 15, 872, 1914; "Further Remarks on Electron Vapor Problem," Physik Zeit, 20, 220, 1919
   Schottky, W., "On Electrons Emerging from Glowing Wires at Retarding Voltages," Ann. Physik, 44, 1011, 1914; "On Cold and Warm Electron Discharges," Zeits F. Physik, 14, 63, 1923
10. Dushman, S., "Thermionic Emission," Rev. Modern Phys. 2, 381, 1930
11. Wehnelt, A. and F. Jentsch, "On the Energy of Thermionic Electron Emission, "Ann. D. Physik, 28, 537, 1909
12. Cooke, H. L. and O. W. Richardson, "Absorption of Heat Produced by the Emission of Ions from Hot Bodies," Phil. Mag., 26, 472, 1913; "The Absorption of Heat Produced by the Emission of Ions from Hot Bodies," Phil. Mag., 25, 624, 1913
13. Lester, H. H., "The Determination of the Work-Function When an Electron Escapes from the Surface of a Hot Body," Phil. Mag., 31, 197, 1916
14. Langmuir, I., "The Effect of Space Charge and Residual Gases on Thermionic Currents in High Vacuum," Phys. Rev., 2, 450, 1913
    Langmuir, I., "Thermionic Currents in High Vacuum"; "The Electron Emission from Tungsten and the Effect of Residual Gases," Physik Zeits, 15, 516, 1914
15. Kingdon, K. H. and I. Langmuir, "The Removal of Thorium from the Surface of a Thoriated Tungsten Filament by Positive Ion Bombardment," Phys. Rev., 24, 148, 1923
16. Kingdon, K. H., "Electron Emission from Adsorbed Films on Tungsten," Phys. Rev., 24, 510, 1924
17. Spanner, H. J., "On Thermal Emission of Charged Particles," Ann. Physik, 75, 609, 1924
18. Zwikker, C., "Thermionic Emission of Metals — Tungsten, Molybdenum, Thorium, Zirconium and Hafnium," Proc. Acad. Sci. Amst., 29, 792, 1926
19. Dushman, S., and J. W. Ewald, "Electron Emission from Thoriated Tungsten," Phys. Rev., 29, 857, 1927
20. DuBridge, L. A., "The Thermionic Emission from Clean Platinum," Phys. Rev., 32, 961, 1928
21. Jones, H. A. and I. Langmuir, "The Characteristics of Tungsten Filaments as Functions of Temperature," Gen. Elec. Rev., 30, Part I, p 310-319, Part II, p 354-361, Part III, p 408-412, 1927
22. Ayer, R. B., "Use of Thoriated-Tungsten Filaments in High Power Transmitting Tubes," Proc. I R E , 40, 591, 1952
23. Ayer, R. B., "High-Power Industrial Vacuum Tubes Having Thoriated-Tungsten Filaments," Trans. A.I.E.E., 72, pt. 1, 121, 1953
24. Langmuir, I., "The Electron Emission from Tungsten Filaments Containing Thorium," Phys. Rev., 4, 544, 1914
25. Koller, L. R., The Physics of Electron Tubes, 1st ed., McGraw-Hill, New York, 1934
26. Parker, W. N., and M. V. Hoover, "Gas Tubes Protect High Power Transmitters," Electronics, 29, p 144, January, 1956
27. Horsting, C. W., "Carbide Structures in Carburized Thoriated-Tungsten Filaments," J. Applied Phys., 18, 95, 1947
28. Code B1
29. Roeser & Wensel, "Properties of Tungsten," Handbook of Chemistry & Physics, 51st Edition, Chemical Rubber, 1949
30. Code B2
31. Advances in Electronics, Vol. 5, p 170-210, Academic Press, 1953

Photoelectric and Secondary-Electron Emission

A. H. Sommer

Princeton

In the past, progress in the field of photoelectric and secondary-electron emission was made mainly on an empirical basis. In recent years, however, general understanding of the solid state has advanced sufficiently to provide some qualitative ideas concerning the mechanisms of photoelectric and secondary-electron emission. Such understanding is more advanced in the field of photoemission than in that of secondary-electron emission. Therefore, photoelectric emission will be treated in considerable detail, whereas the discussion of secondary-electron emission will be confined essentially to experimental results.

PHOTOELECTRIC EMISSION

GENERAL CONSIDERATIONS

The basic law of photoelectric emission is the Einstein equation

$$eV = h\nu \tag{1}$$

according to which a light quantum or photon $h\nu$ is converted into energy eV. The energy, expressed in electron-volts, which the electron obtains from the photon is given by

$$V = \frac{h\nu}{e} = \frac{hc}{e} \cdot \frac{1}{\lambda} \tag{2}$$

Where h is Planck's constant, V, λ, and c are, respectively, the frequency, wavelength, and velocity of the light; and e is the electron charge. Substitution, in Eq. (2), of the numerical values for h, c, and e gives

$$V \approx \frac{\lambda}{12{,}000} \tag{3}$$

where λ is in angstroms. This equation means that for visible radiation, covering approximately the range from 4000 to 8000 angstroms, photoelectrons have an energy of between 1.5 and 3 electron-volts; however, this value is not the energy with which the electrons can escape into the vacuum, because part of it is needed to overcome the surface barrier. Thus, the maximum energy with which the electron is emitted is reduced by the value of the work function ϕ and, therefore, Eq. (3) becomes

$$V = \frac{12{,}000}{\lambda} - \phi \tag{4}$$

Equation (4) shows that there exists a long wavelength threshold λ_0 for photoemission given by

$$\lambda_0 = \frac{12{,}000}{\phi} \tag{5}$$

Electrons cannot be emitted if the wavelength of the light exceeds λ_0. To respond to visible radiation, therefore, the photoemissive material must have a ϕ lower than 3.0 and preferably as low as 1.5 electron-volts. This statement is strictly applicable only for metals; the case of semiconductors will be discussed later.

A low value of ϕ is an essential condition for photoemission*, but for a more detailed understanding of emissive materials, it is useful to consider photoemission as a three-step process. The first step is the absorption of light and its conversion into electronic energy. The second is the motion of the electron through the solid to the surface, and the third is the escape of the electron through the surface barrier. With these three steps in mind, it will be shown that the efficiency of photoemission differs considerably for the two basic types of material, namely, metals which contain a large number of free electrons and semiconductors or insulators which, for practical purposes, do not contain free electrons.

METALLIC PHOTOEMITTERS

In the case of metals, a large amount of incident light is reflected and, therefore, lost for photoemission. If the free electrons in a metal were really completely free, the law of conservation of momentum would prevent the absorption of photons altogether. (In reality, there is no completely free electron in a solid because some residual binding forces are exerted by the excess positive charge and, therefore, some light is absorbed by the metal.) Thus, it is apparent that the first step in the process of photoemission, i.e., the absorption of light, is of low efficiency in metals.

The second step is the motion of the electron through

*This statement applies, of course, only to visible radiation. The importance of ϕ decreases with increasing photon energy, i.e., for radiation in the far-ultraviolet and X-ray range.

the solid. Because of the large number of free electrons in a metal, electrons produced at a distance from the vacuum interface are likely to suffer many collisions with other free electrons (electron scattering). The energy loss associated with the scattering process reduces the probability of their reaching the surface with sufficient energy to overcome the surface barrier. Therefore, in practice, only light absorbed very close to the surface is likely to be effective for photoelectric emission.

Finally, in the third step, the electron has to escape through the surface barrier, i. e., the metal must have a sufficiently low work function. According to Eq. (5), the work function ϕ of metals can be determined from the threshold wavelength of photoelectric emission. It may be mentioned in passing that the photoelectric and thermionic work functions of metals are identical so that ϕ can also be determined by measurement of the change of thermionic emission with temperature (Richardson Plot). The only metallic elements having a low work function are the alkali metals and, to a lesser degree, the alkaline earth metals (see Table I).

Table I

Metal	Work Function* (ϕ)—electron-volts	Metal	Work Function* (ϕ)—electron-volts
Beryllium (Be)	3.9	Lithium (Li)	2.5
Magnesium(Mg)	3.7	Sodium (Na)	2.3
Calcium (Ca)	2.7	Potassium (K)	2.2
Strontium (Sr)	2.7	Rubidium (Rb)	2.1
Barium (Ba)	2.5	Cesium (Cs)	1.9

*Experimental difficulties make all ϕ values doubtful within wide limits.

The work functions of the alkali metals decrease in the series lithium→sodium→potassium→rubidium→cesium, hence cesium has a lower work function than any other chemical element. Without going into detail, it may be mentioned that there is a qualitative correspondence between work function and ionization energy. The alkali metals have low ionization energy and work function because the single electron in the outermost orbit of the atom is relatively loosely bound.

Cesium, although it is the most sensitive of the elementary photoemitters in the visible light range, has a quantum efficiency of only approximately 0.01 per cent, i. e., 10,000 photons are needed to produce one photoelectron. The reasons for this low quantum efficiency have been discussed above; they are, predominantly, large light reflection, low light absorption, and loss of electrons by electron scattering within the material. With the development of the composite materials, to be described later, metallic emitters have lost practical importance for the detection of visible radiation. They are still occasionally used in the ultraviolet region because metals having high work functions may be advantageous for applications where sensitivity to visible radiation is not wanted.

SEMICONDUCTING PHOTOEMITTERS

In semiconductors and insulators, as shown in the conventional energy-band model of Fig. 1, the edge of the filled (valence) band is separated by a forbidden band E_g from the edge of the conduction band. Insulators and semiconductors differ merely in that the forbidden band is larger for insulators.

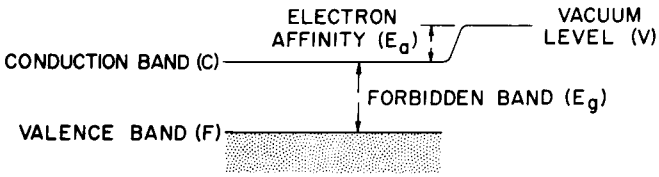


Figure 1. Conventional Energy-Band Model of a Semiconductor

In order that a semiconductor can absorb light, the energy of the photon has to be large enough to raise an electron from the valence band into the conduction band. Therefore, the long wavelength threshold of light absorption will correspond to the energy E_g. An electron raised to the conduction band may produce photoconductivity, but, in order to escape as a photoelectron into the vacuum, it has to have additional energy to reach the vacuum level V. The energy difference between conduction band and vacuum level is called electron affinity E_a. Thus, to produce photoemission from a semiconductor, a photon must have the minimum total energy (E_g+E_a). The practical importance of semiconductors for photoemission lies in the fact that materials have been found having (E_g+E_a) values smaller than the ϕ of any metal. The foregoing analysis applies to semiconductors of the intrinsic type. In "impurity" semiconductors, which will not be considered in detail, photons of lower energy could release some electrons from donor levels within the forbidden band, but the efficiency of this effect would be very low because of the small concentration, and hence small light absorption, of impurities in the effective region near the surface.

The measurement of E_g and (E_g+E_a) is, in principle, not difficult. E_g can be determined from the long wavelength threshold of the light absorption or from the long wavelength threshold of photoconductivity and, in some cases, from the change of resistance with temperature; (E_g+E_a) is derived from the threshold of photoemission. Thus, the threshold for photoemission should always be at shorter wavelengths than that for photoconductivity and light absorption. Actual measurements of these values will be referred to later.

Conditions for the motion of an electron through the solid are also more favorable for semiconductors than for metals. There are practically no free electrons in the semiconductor and, therefore, the mean free path

of an electron released by light is much greater than in a metal. Although the electron loses energy by collisions with the lattice in a semiconductor, lattice scattering is a much less efficient process than electron scattering. As a result, electrons can be released from a depth of more than 200 angstroms below the vacuum interface.

Finally, the escape of the electron through the surface barrier of the semiconductor is determined by the electron affinity. It is worth mentioning that, although the work function ϕ of metals is fairly accurately known, the values for E_a can only be determined as the difference of $(E_g+E_a)-E_g$. As has been explained, these energy values are derived from threshold measurements which are inherently not precise because they depend on the sensitivity of the measuring equipment. Thus, electron affinity values are even less accurate than the ϕ values of metals.

As a result of the more complete light absorption and the greater depth from which electrons can be released, semiconducting photoemitters have proved much more efficient than metallic emitters. In the following, semiconductor photocathodes, particularly those which have found practical use, are described in greater detail.

PRACTICAL PHOTOCATHODES

Photocathodes can be formed as opaque layers with the light incident on the vacuum interface or as semitransparent layers with the light usually incident on the interface between cathode and (transparent) supporting material. Opaque cathodes can be formed in arbitrary thickness, either directly on the glass wall of an evacuated tube or on a conducting base material such as a solid metal electrode or a conducting coating on glass or mica. For semitransparent cathodes, the thickness is critical because the layer must be thick enough to absorb most of the incident light, but thin enough for photoelectrons originating near the cathode-glass interface to have a high probability of escape. The optimum thickness is usually in the range of 100 atomic layers.

Despite the more critical activation process, semitransparent cathodes are preferred for devices such as multiplier phototubes and television camera tubes because they simplify light optics and electron optics. Contrary to expectation, it has been found that light incident on the glass interface of a semitransparent cathode may release more photoelectrons than light incident on the vacuum interface. This effect has been attributed to the complex mechanism of light absorption when the wavelength of the light is greater than the film thickness.

Photoemitters having a useful quantum efficiency have several features in common: First, as has already been mentioned, they are semiconductors. Second, they all contain alkali metals. Third, in all compounds containing a single alkali metal the threshold wavelength increases in the series lithium → sodium → potassium rubidium → cesium so that the cesium compound is always the most sensitive to visible radiation. The following summary of the forming process, photoelectric characteristics, and other properties of photocathodes is confined, therefore, in most cases to the cesium compound and to semitransparent cathodes because the activation process and the properties are basically similar for compounds with other alkali metals and for opaque cathodes. The description of individual photocathodes will proceed from the simpler to the more complex materials rather than in historical sequence. For more details on the formation of photocathodes, the reader is referred to Refs. 1, 2 and 3.

The Antimony-Cesium (Sb-Cs) Cathode

Since 1936[4] many photoemissive materials have been found which have the common characteristic that they are semiconducting intermetallic compounds of alkali metals with metals of the fourth, fifth, or sixth group of the Periodic System. The Sb-Cs cathode is outstanding among these materials because it has high photosensitivity, is simple to produce, and has been the subject of numerous experimental studies aimed at a better understanding of the photoelectric mechanism.

Formation. The formation of an Sb-Cs cathode (as well as that of all subsequent cathodes) requires a vacuum in the range of 10^{-6} millimeters of mercury or less. The first step in the process is the evaporation of an Sb layer on the supporting material. The thickness of the layer is monitored by measurement of the transmission of white light through the layer. The evaporation is stopped when the transmission has decreased about 20 per cent. The tube is then heated to between 120 and 150 C and the layer is exposed to Cs vapor. Because alkali metals oxidize at once upon exposure to air, the Cs required for the activation is usually produced within the evacuated tube itself or in an attached side tube. A convenient technique is the use of a nickel pellet filled with an intimate mixture of a Cs salt and a reducing agent (for instance, one part of cesium chromate powder with four parts of pure silicon or zirconium[5] powder). Cesium is evolved from this mixture when it is heated to temperatures greater than 700 C.

When the tube and the Sb layer are at a temperature of 120 to 150 C, the Cs readily reacts with Sb to form the photosensitive compound. This reaction is accompanied by three characteristic phenomena. First, the Sb layer loses its metallic appearance and assumes a red color in transmitted light. Second, the resistance of the layer rises rapidly by many orders of magnitude above that of the original Sb layer, indicating the formation of a semiconductor. Third, white light produces photoelectric emission which rises to a peak and then drops rapidly with continued addition of Cs. During the activation process, the photoemission is usually monitored. When the photoemission passes over the peak, the introduction of Cs is stopped and the baking process is then continued until the photocurrent reaches a constant peak value. After the tube is cooled, the sensitivity, particularly for red light, can be further improved by careful exposure of the cathode to oxygen until the photoemission again reaches a peak. Oxygen is conveniently introduced by heating, in a side-tube, a salt such as potassium permanganate ($KMnO_4$) or potassium chlorate ($KClO_3$) to the temperature at which it releases oxygen. As an alternative, the oxygen may be introduced by a thin layer of manganese oxide on

which the original Sb layer is deposited.

Properties. Typical response curves of Sb-Cs cathodes before and after surface oxidation are shown in Fig. 2. The most important features of the oxidized cathode are the high quantum efficiency of up to 20 per cent at wavelengths below approximately 4200 angstroms and a practical threshold wavelength of approximately 7000 angstroms. For practical purposes the sensitivity of a phototube is usually given in microamperes per lumen. The lumen used for this measure of sensitivity is based on the color distribution of a tungsten lamp at 2870 K. Phototubes having Sb-Cs cathodes usually have a sensitivity of 40 to 80 microamperes per lumen, and occasionally as high as 100 microamperes per lumen.

In the detection of very low light levels, the thermionic emission of a cathode is an undesirable feature becuase it may limit the signal-to-noise ratio. The thermionic emission of the Sb-Cs cathode is usually below 10^{-4} amperes per square centimeter at room temperature.

The reproducible performance of the Sb-Cs photocathode indicated that the material should have a well-defined chemical composition. Quantitative studies[6] suggested the stoichiometric formula Cs_3Sb. However, with present analytical methods, it is not possible to detect minute excess amounts of Sb or Cs within the cathode material or adsorbed to the surface or both. It may well be that just such a small excess of one of the components is of significance in the photoelectric mechanism.

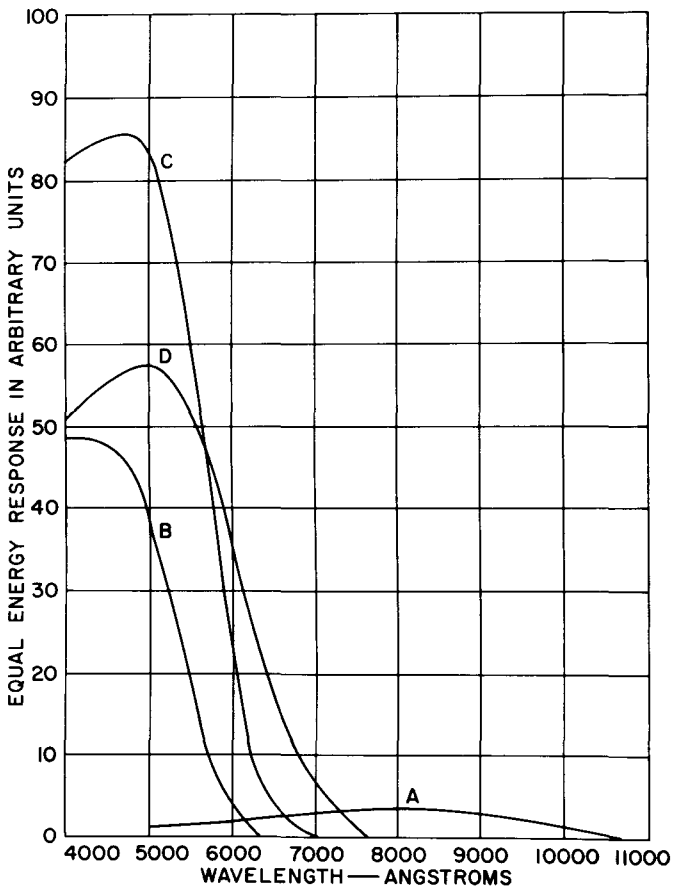


Figure 2. Spectral Sensitivity Characteristics of Various Photoemissive Materials Used as Cathodes in Phototubes: curve A, Ag—O—Cs; curve B, Sb—Cs; curve C, Sb—Cs—(O)

Values of the band-gap energy E_g (see Fig. 1) have been reported[7] as determined from measurements of light absorption, photoconductivity, and curves of log σ vs 1/T where σ is the conductivity and T is the absolute temperature. The electron affinity E_a had been computed from photoemission thresholds; attempts have also been made to measure the thermionic work function. Finally, measurements of the thermoelectric[8] and Hall[9] effect have been made to establish the sign of the current carriers in Cs_3Sb; at present, the evidence is in favor of p-type conduction[10].

Miscellaneous Cathodes of the Sb-Cs Type

There are other photoemitters which are semiconducting stoichiometric compounds of an alkali metal with another metal, but differ from Cs_3Sb by having lower sensitivity in the visible spectrum. The Sb may be replaced by metals of the fourth group such as germanium, tin, or lead (Ge, Sn, Pb), the fifth group such as arsenic or bismuth (As, Bi), and the sixth group such as selenium or tellurium (Se, Te). Two of these compounds deserve special mention: Cs_3Bi [4] because, among the cathodes not containing Sb, it comes closest to Cs_3Sb in sensitivity to visible radiation although its quantum efficiency is down by a factor of about ten; and Cs_2Te because it has very high quantum efficiency in the ultraviolet (up to 30 per cent)[11]. As mentioned in the introduction, photocathodes may also be formed by replacing Cs with other alkali metals. The response curves of K_3Sb and Na_3Sb are shown in Fig. 3.

Multi-Alkali Cathodes

Although Cs is superior to other alkali metals in all types of cathodes, it has been found[12] that the combination of two or more alkali metals with antimony may in some cases produce higher sensitivity than Sb with any one of these alkali metals. This "multi-alkali effect" has been definitely established for two cathodes of the general formulae Sb-K-Na and Sb-K-Na-Cs. The formation process for these cathodes is essentially similar to that of the Cs_3Sb cathode, but the introduction of the alkali metals has to be carefully controlled because the ratio in the final layer is critical.

The spectral response curve for Sb-K-Na is shown in Fig. 3; comparison with the curves for the corresponding single-alkali cathodes K_3Sb and Na_3Sb shows the magnitude of the multi-alkali effect. The sensitivity of the Sb-K-Na cathode is similar to that of superficially oxidized Cs_3Sb but it has the advantage of much lower thermionic emission at room temperature: less than 10^{-16} amperes per square centimeter.

The response curve of the Sb-K-Na-Cs cathode is shown in Fig. 3 and covers the visible spectrum more effectively than any other known photoemitter. The sensitivity to tungsten light often exceeds 200 microamperes per lumen. The thermionic emission at room temperature, despite the longer threshold wavelength,

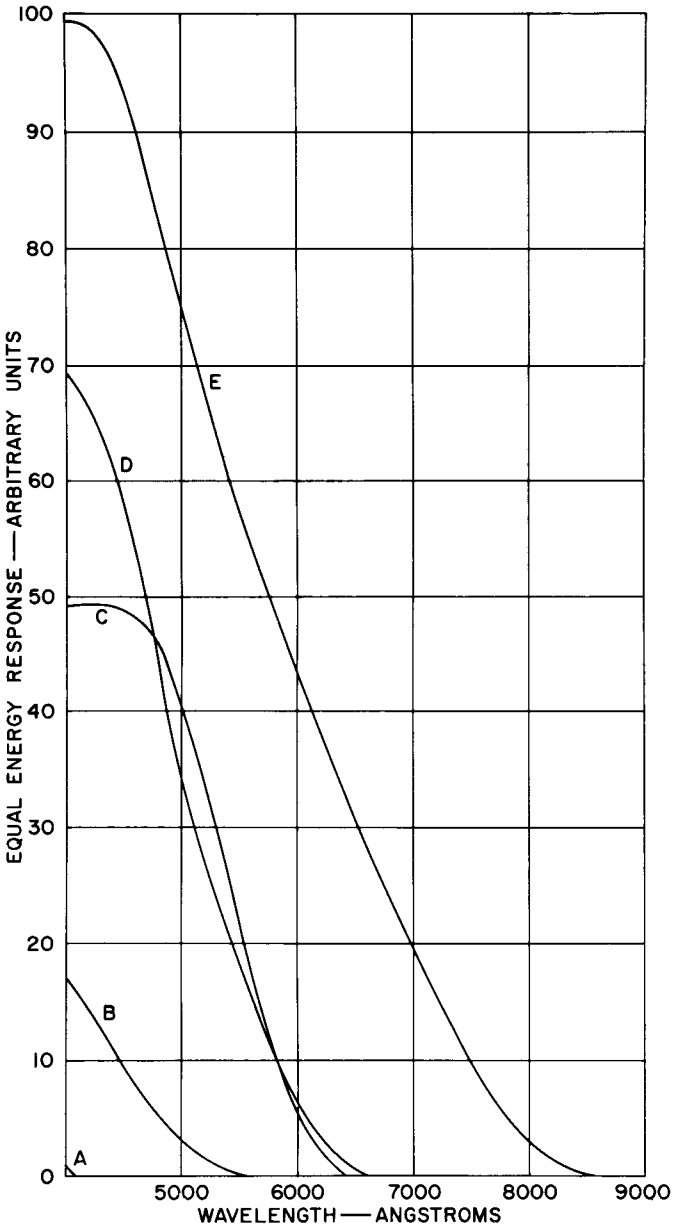


Figure 3. Spectral Sensitivity Characteristics of Photoemissive Materials of the Alkali Antimonide Type: curve A, Sb—Na; curve B, Sb—K; curve C, Sb—Cs; Curve D, Sb—K—Na; curve E, Sb—K—Na—Cs

is smaller than that of superficially oxidized Cs_3Sb, usually about 10^{-16} amperes per square centimeter.

There is no explanation for the "multi-alkali effect," but measurements of light absorption, photoconductivity, and chemical composition make it certain that the Sb-K-Na-Cs cathode is basically an Sb-K-Na cathode whose electron affinity is reduced by a surface layer of Cs. A systematic study[13] has shown that the improvement in red response of the Sb-K-Na-Cs cathode over that of the Cs_3Sb cathode is caused by a lower value of the band-gap energy, while the high quantum efficiency of the Sb-K-Na cathode, compared with that of K_3Sb and Na_3Sb, is associated with a reduction of the electron affinity.

The Silver-Oxygen-Cesium (Ag-O-Cs) Cathode

Formation. The first step in the preparation of an Ag-O-Cs cathode is the formation of a silver base. The base can be produced in a number of ways, such as electroplating, chemical deposition, or evaporation. The last method is preferable because it produces the cleanest and most uniform surface. It also is the only method by which a semitransparent cathode of controlled thickness can be made. The silver base is oxidized by exposure to a glow discharge in 0.1- to 1-millimeter pressure of oxygen. The oxidation is accompanied by a characteristic color change which allows accurate control of the process. The colors go through golden-yellow to blue and then green. The best results are obtained when the oxidation is stopped at the blue color. (This procedure applies to opaque cathodes; semitransparent silver layers are completely oxidized and have almost 100 per cent transparency.)

After the tube has been evacuated again, the silver oxide is exposed to Cs vapor and at about 150 C a reaction takes place. As soon as the photoemission passes through a maximum, the introduction of Cs is stopped and the tube is kept at the indicated temperature until the photocurrent reaches a stable maximum value. It has been found[14] that higher sensitivity is obtained by the additional evaporation of a small amount of silver, followed by another baking until maximum photocurrent is obtained.

Properties. A typical color response curve for an Ag-O-Cs cathode is shown in Fig. 2. The maximum quantum efficiency in the visible and infrared regions is only 0.5 per cent at 8500 angstroms, but, because the threshold wavelength is beyond 12,000 angstroms, this cathode is valuable for infrared work; in fact, this photoemitter is the only one known having a usable infrared response. The tungsten light sensitivity is 30 to 50 microamperes per lumen. The thermionic emission at room temperature varies over a wide range (10^{-13} to 10^{-9} amperes per square centimeter) even among cathodes having similar photoelectric behavior; 10^{-12} amperes per square centimeter can be considered an average value.

Despite a great amount of experimental work, the mechanism of the Ag-O-Cs cathode is little understood. It is known only that the reaction of silver oxide with Cs leads to cesium oxide mixed with Ag, but there is no explanation for the low quantum efficiency nor for the fact that the silver apparently cannot be replaced by another metal.

The Bismuth-Silver-Oxygen-Cesium (Bi-Ag-O-Cs) Cathode

The absence of red response in the Cs_3Sb cathode and the low efficiency of the Ag-O-Cs cathode in the visible range led to unsuccessful attempts at combining these two materials. However, it was found (see Ref. 3, p. 300) that if Sb in this combination is replaced by Bi, a cathode can be produced which approaches in its properties the desired compromise.

Formation. The first step of the process consists of

the evaporation of a thin Bi layer having a light transmission of about 60 per cent. This is followed by the evaporation of Ag until the transmission is reduced to about 50 per cent. This bismuth-silver layer is then oxidized by a discharge in oxygen or simply by exposure to oxygen until the transmission has increased again to about 55 per cent. Finally, the layer is activated with Cs at approximately 150 C until maximum sensitivity is obtained.

Properties. A typical color response curve is shown in Fig. 2; although the peak quantum efficiency is lower than that of Cs_3Sb, the drop toward longer wavelengths is more gradual so that the whole visible range is covered. The total sensitivity is usually 30 to 60 microamperes per lumen and sometimes is as high as 90 microamperes per lumen. The cathode is particularly suited for applications where panchromatic response is required. The thermionic emission at room temperature is approximately 10^{-13} amperes per square centimeter, which is about ten times that of Cs_3Sb.

Little is known about the mechanism of this very complex photoemitter. It is particularly puzzling why Bi in this combination is better than Sb, by contrast with the superiority of the Cs_3Sb cathode over the Cs_3Bi cathode. As in the case of the Ag-O-Cs cathode, there is no explanation for the function of the Ag.

The most important characteristics of the described photocathodes are summarized in Table II. It must be emphasized that the values for the wavelength of peak response (λ peak) and threshold response (λ_0) and, particularly, the values for thermionic emission vary considerably from tube to tube.

Matching of Photocathode and Light Source

Of the above described cathodes, the Cs_3Sb cathode is the easiest to produce and, therefore, it is used in all general-purpose tubes where high sensitivity to white light is required, However, in some applications other cathodes may be preferable. As an indication of the variety of requirements, a few specific examples follow:

1. In scintillation counters, the light source is a phosphor which emits chiefly in the blue region of the spectrum. Hence, the Cs_3Sb cathode is most suitable, although the lower thermionic emission of the Sb-K-Na cathode may render the latter superior for the detection of extremely low signals.
2. In television camera tubes, particularly those for color television, panchromatic response is desirable. Therefore, the Bi-Ag-O-Cs cathode is generally used. However, the more recently developed K-Na-Cs-Sb cathode is gaining in importance because of its higher sensitivity, particularly to red

Table II

Characteristics of Photocathodes

Photocathode Material	Phototube Response Number*	Wavelength of Peak Response λ (peak) angstroms	Maximum Quantum Efficiency at λ (peak) per cent	Wavelength of Threshold Response † λ_0 angstroms	Maximum Luminous Sensitivity microamperes per lumen	Average Thermionic Emission at Room Temperature—amperes per square centimeter
Sb-Cs(O) (opaque)	S-4	4000	25	7000	120	10^{-15}
Sb-Cs	---	4500	10	6500	25	$< 10^{-14}$
Sb-Cs(O) (semitransparent)	S-11	4500	20	7000	90	10^{-15}
Ag-O-Cs	S-1	8500	0.8	12000	50	10^{-12}
Bi-Ag-O-Cs	S-10	4500	10	7500	90	10^{-13}
Sb-K	---	3500	---	---	5	---
Sb-Na	---	2700	---	---	< 0.1	---
Sb-K-Na	---	~ 3700	> 25	6200	60	$< 10^{-16}$
Sb-K-Na-Cs	S-20	~ 4500	> 30	8500	230	10^{-16}

*Includes effect of envelope material.

† Response at threshold wavelength is one per cent of response at peak wavelength.

light. Panchromatic response is also advantageous for many scientific applications such as spectroscopy, astrophysics, and biology.

3. For the detection of infrared radiation, only the Ag-O-Cs cathode is available. The image converter tube may be cited as an important application of this material.

SECONDARY-ELECTRON EMISSION

GENERAL CONSIDERATIONS[15]

Secondary emission and photoemission have many basic features in common; the main difference between the two types of energy conversion is that the energy of bombarding photons at visible radiation frequencies is less than three electron volts, whereas the energy of bombarding primary electrons can be varied at will and is for practical purposes usually two orders of magnitude larger than that of visible-radiation photons.

With this quantitative difference in mind, secondary-electron emission can also be considered as a three-step process similar to that described previously for photoemission. The first step, the absorption of a primary high-energy electron and its conversion into one or more low-energy secondary electrons, is still little understood, both for metals and for semiconductors. The second step, the motion of the secondary electron through the solid, does not differ substantially from the corresponding step in the case of photoemission. Again, electron scattering reduces the depth from which electrons can reach the surface and, thus, conditions are again more favorable for semiconductors than for metals. The third step, the penetration of the surface barrier, is again determined by the work function for metals and by the electron affinity for semiconductors. By comparison with photoemitters, there is a difference in degree inasmuch as the average energy of secondary electrons is greater than that of photoelectrons so that secondary electrons can be released from metals having a higher work function or from semiconductors having a larger band gap.

PARAMETERS OF SECONDARY-ELECTRON EMISSION

Whereas photoemitters are defined by the parameters of spectral response and quantum efficiency, the three important parameters of secondary-electron emitting materials are: (1) gain factor δ as a function of primary electron energy, (2) energy distribution of emitted secondary electrons, and (3) gain factor δ as a function of angle of incidence of primary electrons. Before specific materials are described, a few general remarks on these parameters may be useful.

Gain Factor δ as a Function of Primary Energy

All secondary-electron emitters — metals and semiconductors alike — show the same general shape for the curve of the gain as a function of primary energy (see Fig. 4). With increasing primary energy, δ rises and reaches a peak in the 500- to 1000-volt range, followed by a very gradual decrease. This behavior is explained as follows: On the one hand, with increasing primary energy an increasing number of secondaries is produced. On the other hand, more energetic primary electrons will, on the average, penetrate to a greater depth so that at excessive primary energy too many secondary electrons will be released at a depth from which they cannot reach the surface. Thus the peak value of δ represents the point at which the increase in the production of secondary electrons is compensated by the decrease of the number reaching the surface. This point will be reached at lower δ values for metals than for semiconductors as a result of electron scattering.

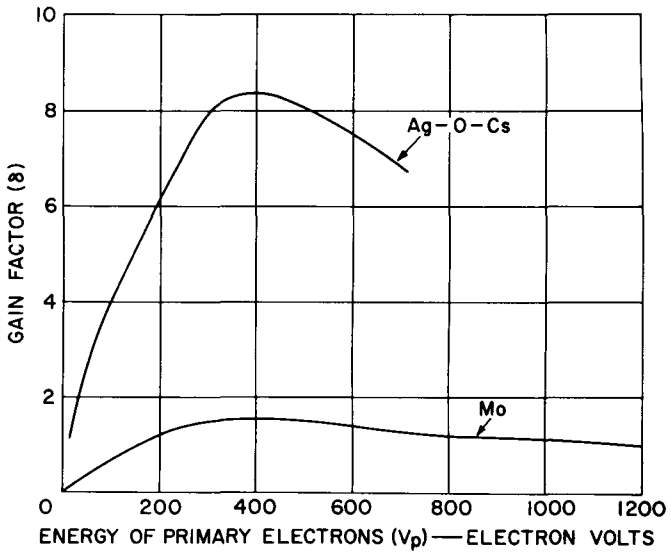


Figure 4. Secondary Emission as a Function of Primary-Electron Energy

Energy Distribution of Secondary Electrons

Whereas the energy distribution of photoelectrons is limited to the very narrow range from zero to $(h\nu-\phi)$ (i. e., to considerably less than three electron volts for visible radiation), the energy spectrum of secondary electrons extends from zero up to the energy of the primaries. This effect is not only of academic interest but also of practical importance because the large spread of initial velocities affects the electron optical properties of tubes employing secondary-electron emission.

A typical distribution of secondary-electron energies is shown in Fig. 5. This curve can be qualitatively interpreted as follows: The peak at the energy value of the primary electrons represents elastically reflected primary electrons. The small number of secondary electrons emitted in the intermediate range are inelastically reflected primary electrons. Finally, the peak at low energies is produced by the "true" secondary electrons. This peak is in the range from 1 to 4 electron-volts, which is considerably larger than that for photoelectrons.

Gain Factor δ as a Function of Angle of Incidence

It has been found that for primary electrons having energies exceeding about 100 electron-volts, δ in-

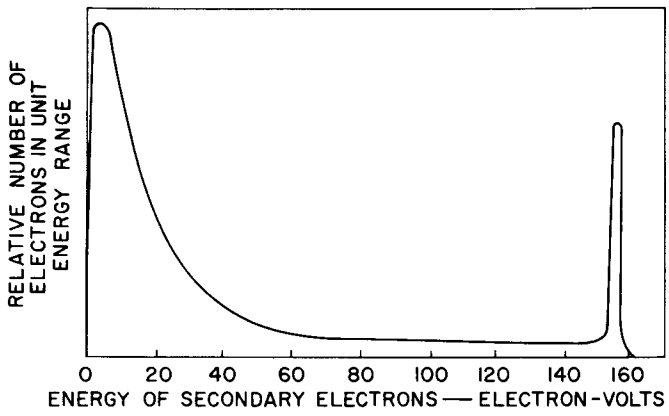


Figure 5. Velocity Distribution of Secondary Electrons

creases as the angle of incidence deviates from the normal. This effect is readily explained by the fact that primary electrons penetrating into the material to a certain depth will release the secondary electrons closer to the surface if the incidence is glancing rather than if it is normal. The effect is not noticeable for primary electrons below 100 electron-volts because slow primary electrons are absorbed so close to the surface that a change of angle would have a negligible effect. For practical purposes, the angle of incidence is not very important because in most applications primary electrons having energies of less than 200 to 300 electron volts are used.

PRACTICAL SECONDARY EMITTERS

As has been pointed out, semiconductors are more efficient secondary emitters than metals and, therefore, are generally used for such applications as dynodes in multiplier phototubes. However, it must be mentioned that there is also a practical need for surfaces of exceptionally low δ, for instance, for grids and other electrodes in many types of tubes. For such applications, graphite is often used because it has a peak gain close to unity.

This article is mainly concerned with secondary emitters having high values of δ. From the earlier discussion, it is understandable that all efficient photoemitting materials are also efficient as secondary emitters. In addition, some materials whose large band-gap energy E_g prevents photosensitivity to visible radiation have found use as secondary emitters.

Secondary-Emitting Photocathodes

Both the Ag-O-Cs and the Cs_3Sb photocathodes have high δ as secondary emitters, with peaks of 10 and more. They have found extensive use as dynodes in multiplier phototubes because it is advantageous in manufacture if photocathode and dynodes can be activated by the same process. However, Ag-O-Cs dynodes have gradually fallen into disuse because the thermionic emission of Ag-O-Cs surfaces causes additional dark current.

The other photocathodes described above also have a high δ but, because of the complex activation process of these materials, they have not found practical application.

Silver-Magnesium-Oxygen (Ag-Mg-O)

Among the materials which are not photosensitive to visible radiation, the highest gain factor is obtained with partially oxidized silver-magnesium alloys.[16] For metallurgical reasons, these alloys can be made only with a maximum Mg content of about two per cent. It is believed that an essential condition for secondary emission is a surface layer of MgO. It is likely that the Ag is needed only to provide sufficient conductivity and that it does not play a specific part in the emission process, in constrast to the role of Ag in the Ag-O-Cs photocathode.

The oxidation process of the Ag-Mg alloy is very critical in its effect on δ. At present it is believed that heating in oxygen leads to diffusion of oxygen into deeper layers and, thus, prevents a sufficient concentration of MgO at the surface. Better results are obtained by oxidation with larger molecules which do not readily diffuse so that Mg, in turn, has to diffuse to the surface to be oxidized. For this reason, a baking process in water vapor[17] at low pressure or in CO_2[18] is generally used.

Ag-Mg-O dynodes have somewhat lower δ than Cs_3Sb dynodes, but they have the advantage of easier manufacture, particularly for multi-electrode tubes where the formation of an Sb layer by evaporation is inconvenient and difficult. In multiplier phototubes it is generally found that the Cs vapor introduced for the cathode activation increases the δ of Ag-Mg-O dynodes. This effect is most likely due to a lowering of the electron affinity by an adsorbed Cs surface layer.

Copper-Beryllium-Oxygen (Cu-Be-O)

A combination basically similar to Ag-Mg-O, namely, a partially oxidized alloy of Cu and Be, has also found use as dynode material[19]. Again, the maximum amount of Be that can be alloyed with Cu is only about two per cent, and again the oxidation process is critical[20], probably because a surface layer of BeO is essential. The peak of δ for Cu-Be-O is reached at a higher voltage than for Ag-Mg-O so that, for a given voltage below the peak, the efficiency is lower. In the important range of primary electron energies below 300 electron volts, Cu-Be-O has δ values similar to those of Ag-Mg-O. Cu-Be-O appears to have the advantage over Ag-Mg-O that it is less liable to deterioration on exposure to air. In addition there are indications that Cu-Be-O dynodes are more stable under electron bombardment.

REFERENCES

1. Zworykin, V. K., and E. G. Ramberg, Photoelectricity and its Application, Wiley, New York, 1949
2. Sommer, A. H., Photoelectric Tubes, Wiley, New York, 1951
3. Methods of Experimental Physics, Vol. 6B, Academic Press, New York and London, 1959, Chapter on "Photoelectric Emission" by Sommer, A. H., and W. E. Spicer, pp. 376-391

4. Gorlich, P., "About Alloy Photocells" Z. f. Phys., 101, 335, 1936
5. de Boer, J. H., "Reduction of Alkali Metal Compounds with Zirconium," Z. anorg. Chem., 191, 113, 1930
6. Sommer, A. H., "Photoelectric Alloys of Alkali Metals," Proc. Phys. Soc., 55, 145, 1943
7. Wallis, G., "Optical and Electrical Measurements of Cesium-Antimony Layers of Different Compositions," Annalen der Physik, 17, 401, 1956 (This paper contains copious references to earlier work.)
8. Sakata, T., "Some Experimental Studies of the Conductivity and Thermo-electromotive Force of Cs_3Sb Photocathodes," J. Phys. Soc., Japan, 8, 125, 272, 1953
9. Sakata, T., "Hall Effect in Cesium Antimonide with Audio Frequency Currents," J. Phys. Soc. Japan, 9, 1030, 1954
10. Sommer, A. H. "N-Type and P-Type Conduction in Alkali Antimonide Photoemitters," J. Appl. Phys., 29, 1568, 1958
11. Taft, E., and L. Apker, "Photoemission from Cesium and Rubidium Tellurides," J. Opt. Soc. Am., 43, 81, 1953
12. Sommer, A. H. "New Photoemissive Cathodes of High Sensitivity," Rev. Scient. Instr., 26, 725, 1955
13. Spicer, W. E., "Photoemissive, Photoconductive and Optical Absorption Studies of Alkali-Antimony Compounds," Phys. Rev., 112, 114, 1958
14. Asao, S., "Behavior of the Foreign Metal Particles in Composite Photo-cathodes," Proc. Phys. Math. Soc. Japan, 22, 448, 1940
15. Dekker, A. J., Solid State Physics, Prentice Hall, Englewood Cliffs, 1958, Chapter 17, p. 418-446
16. Zworykin, V. K., J. E. Ruedy and W. W. Pike, "Silver-Magnesium Alloy as Secondary Electron Emitting Material," J. Appl. Phys., 12, 696, 1941
17. Rappaport, P., "Methods of Processing Silver-Magnesium Secondary Emitters for Electron Tubes," J. Appl. Phys., 25, 288, 1954
18. Wargo, P., B. V. Haxby and W. G. Shepherd, "Preparation and Properties of Thin Film MgO Secondary Emitters," J. Appl. Phys., 27, 1311, 1956
19. Allen, J. S., "An Improved Electron Multiplier Particle Counter," Rev. Scient. Instr., 18, 739, 1947
20. Sommer, A. H., "Activation of Ag-Mg and Cu-Be Dynodes," J. Appl. Phys., 29, 598, 1958

Photoconductors for Camera Tubes

C. W. Rector

Lancaster

A photoconductor can be defined empirically as a material whose electrical conductivity is increased by the action of light. The behavior of photoconductors can be predicted to a limited extent on the basis of solid-state physics. As yet, however, even the most advanced theories are inadequate for a detailed analysis of practical materials. Fortunately, a relatively simple theoretical model of a photoconductor is available which has considerable heuristic value and can serve the novice well as a means of explaining and predicting certain important properties of photoconductors. The theory of the model will be developed in this chapter in sufficient detail to permit its fruitful application to several specific photoconductors. These photoconductors will serve as examples to guide the reader in his own use of the theory.

Photoconductors are used in several solid-state devices; light-sensitive photocells and light amplifier panels are perhaps the two most prominent. The main use of photoconductors in electron tubes, however, is in television camera tubes. Currently, the most important commercial application is as the light-sensitive element in the vidicon. Thus this chapter, for the most part, will consider the use of photoconductors in this camera tube.

The preparation of photosurfaces for the vidicon involves a host of variables which must be carefully controlled in the production of a satisfactory tube. Some of the pitfalls to avoid in the preparation and use of these photosurfaces will be considered. In addition, methods for varying some of the electronic characteristics to a limited extent in order to achieve photosurfaces having special properties will be described.

BASIC THEORY

That the orbital electrons of a single isolated atom have a series of distinct energy states available to them is well known. However, if several of these atoms are brought near each other, the individual energy levels will be perturbed. If the atoms are considered as a group, then each level will split into as many distinct levels as there are atoms in the group. The effect is similar to that obtained by linking a line of otherwise independent pendulums by a chain of weak springs.

This splitting for the hypothetical case of six hydrogen atoms in a line is illustrated in Fig. 1. Electron energy is plotted against the lattice constant d, i.e., against the interatomic spacing. At a very large distance from each other, each atom will be relatively unaffected by the presence of the others and the levels will be effectively coincident. However, when the lattice constant, that is to say d, is made smaller, there begins to be a certain amount of coupling between the atoms which causes a slight splitting of the levels. Note that the higher energy levels (those corresponding to the "outer" or least tightly bound electrons) are affected first. The "inner" or more tightly bound electrons are somewhat shielded by the others and, therefore, the corresponding levels do not split until the atoms are close enough to each other to be quite strongly coupled.

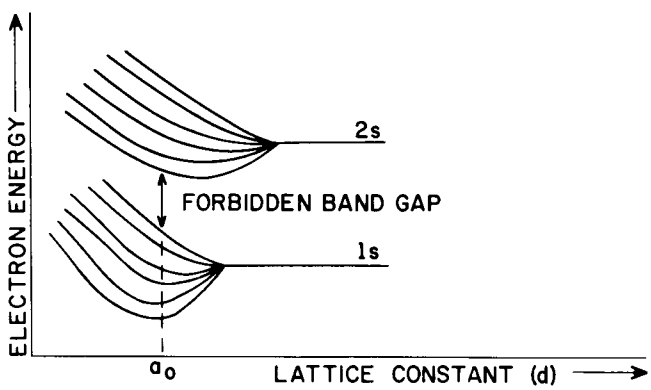


Figure 1. Dependence of Energy Levels on Lattice Constants for Six Hydrogen Atoms

In the case illustrated in Fig. 1, each energy state is split into six distinct states, one for each atom in the group. Were there N atoms in the group, each state would be split into N states. The actual width of the resulting band is not dependent on the number of atoms in the group, but rather on the nature of the atomic energy level from which the band arose, and the degree of coupling: the smaller the lattice constant d the greater the coupling and the wider the resulting band. If a tangible amount of material containing, say, 10^{20} atoms bonded together in a lattice is considered, the individual energy levels are split into 10^{20} distinct levels, forming an energy band the actual width of which, in terms of energy is not very great, so that there exists an extremely high density of energy levels — so dense, in fact, that there is no practical means of distinguishing between the actual energy band composed

of 10^{20} distinct levels and a continuous band, that is to say, a continuum of energy states extending from the bottom of the energy band to its top.

The effect can be pictured more conveniently if a cross section is taken through this graph at the lattice constant a_o and this line smeared out along one of the lattice axes, producing a graph of electron energy versus distance through the lattice. Such a graph is shown in Fig. 2, where the bands are labeled the 1s and 2s allowed bands to correspond with the terminology of Fig. 1. They are termed "allowed" bands because electrons may occupy any energy state in them. Just as the orbital electrons of a single atom can have only those energies corresponding to the various atomic energy levels, 1s, 2s, etc., so, in a crystal lattice, the electrons can occupy only those states within the bands derived from states 1s, 2s, etc. Energies between the allowed bands are denied to the electrons and, as a consequence, this region is termed the "forbidden band gap."

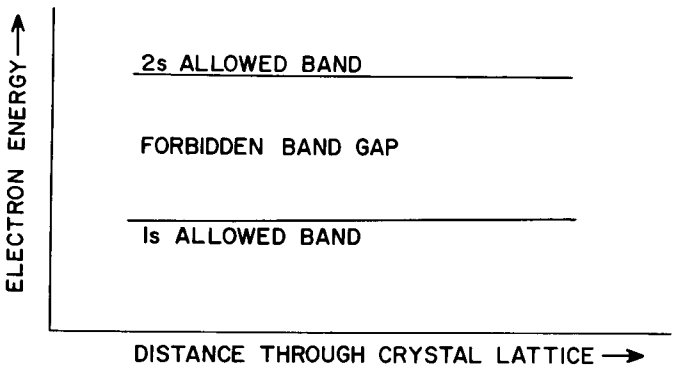


Figure 2. Simplified Band Structure

The inner, more tightly bound (lower energy) electrons play very little part in chemical binding. They are effectively shielded by the outer electrons which are the ones that contribute most to the binding energy in a crystal. These outer electrons are termed the "valence" electrons for they determine the chemical valence of an atom. To fulfill the requirements of minimum energy, and to obey the Pauli exclusion principle which limits to two the number of electrons in any one energy state, the electrons in a solid will occupy the lowest levels in the lowest allowed bands so far as their numbers permit, and will fill them from the bottom upward. At a temperature of absolute zero, when the crystal has its lowest energy, the demarcation level between filled and unfilled states will be sharp. The highest energy electrons, those in the uppermost allowed band which contains electrons at absolute zero, are the valence electrons, and the band in which they lie is termed the "valence" band. The next higher band, the first one which is devoid of electrons at absolute zero, is termed the "conduction" band for reasons to be given later. In the case illustrated in Fig. 2, the two bands do not overlap and there exists a forbidden band gap between them. This case does not apply to all crystals. The energy band diagrams for four different crystal types is presented in Fig. 3. Fig. 3A for instance, has overlapping bands. Such a material will be a good conductor of electricity. Another example of a good conductor is a material with an incompletely filled band shown in 3C. Why it performs this way is of interest. In a solid, the electrons are moving about randomly. The energy and momentum of each electron is specified by the quantum mechanical requirements associated with the particular energy state the electron occupies. Because these motions are random, no net electrical current can result. But if an external electric field is impressed upon the solid, a corresponding force acts upon the electrons which tends to add a drift velocity component to their random movements. However, since this would also add a small increment to their energies, the addition of a drift velocity component can only be realized if there exist available energy states into which, with the small amount of energy available, an electron can be lifted. If a solid has an incompletely filled band, these slightly higher energy states will be available and, therefore, conduction can take place. Or again, if (as in Fig. 3A) the allowed bands overlap, an electron in the top region of the uppermost filled band can move continuously into the next higher band.

In Fig. 3B, however, and in Fig. 2, there exists a wide energy gap between the highest filled band and the lowest empty band. If this highest filled band, the so-called valence band, is completely filled, then for an electron to increase its energy, it must acquire enough energy to be boosted into the next allowed band, which is termed the conduction band. Consider that at room temperature the thermal energy kT* is only about 0.025 electron volt, then a material having a band gap greater than about 2 electron volts (ev) would be very much an insulator, because at normal fields the energy obtainable from the field is appreciably less than kT and only those extremely few electrons in the far reaches of the tail of the thermal energy-distribution curve have energies greater than the 2 ev required to raise an electron from the valence band to the conduction band. Note, however, that should an electron be excited from the valence band to the next-higher normally unfilled band, it would act as a conduction electron, for that band is an incompletely filled band and the electron can, therefore, acquire the small increments of energy offered by the external electric field.

Because the band gap is very wide in an insulator, very few electrons can be thermally excited into the conduction band. If the forbidden band gap is narrow (as in Fig. 3C) an appreciable number of valence electrons may, at room temperature, be thermally excited into the conduction band. The number so excited will be few compared to those available in a material with either an incompletely filled valence band or with overlapping bands. As a result, such solids will not conduct electricity as well as metals. For this reason, they are termed semiconductors. The particular type shown in Fig. 3C is called an intrinsic semiconductor because its conductivity is due to its narrow band gap rather than to impurity content.

Fig. 3 shows a semiconductor with impurity atoms

*k is here the Boltzmann Constant, equal to 1.3805×10^{-6} erg/^{o}K

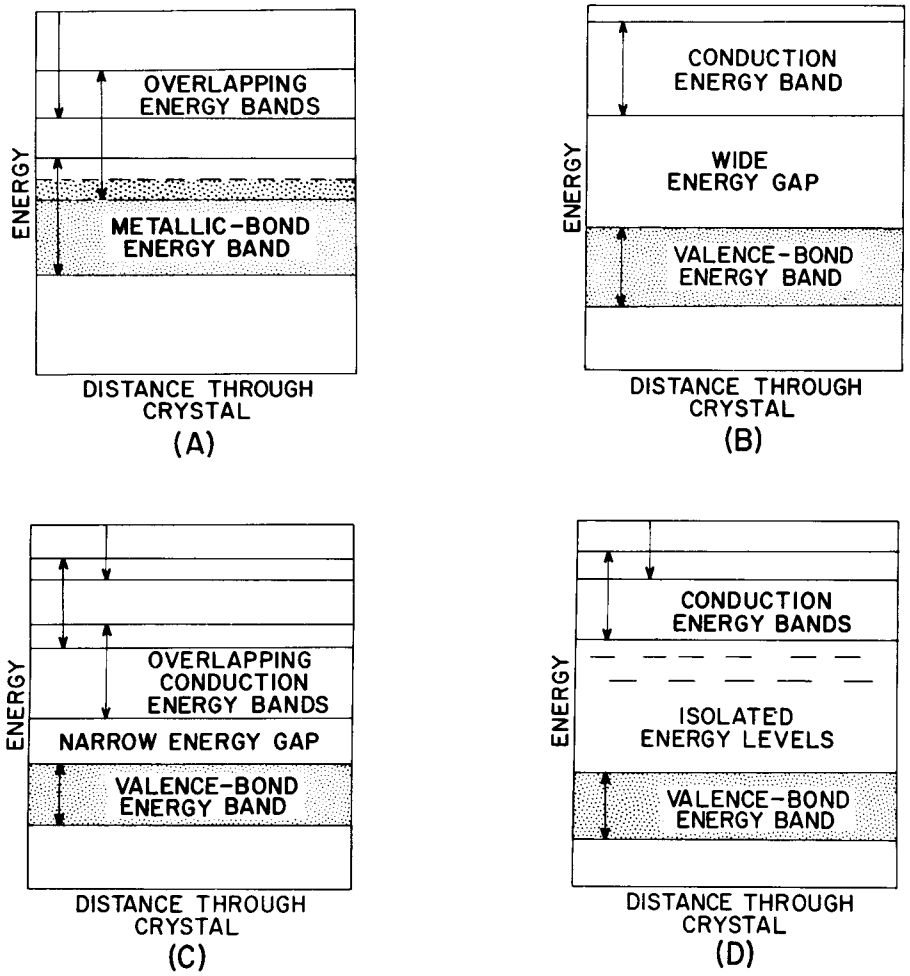


Figure 3. Energy Band Diagrams for Several Crystal Types (after Shockley): (A) Metal, (B) Insulator, (C) Intrinsic Semiconductor, (D) Impurity Semiconductor

of the kind which have, after bond requirements are satisfied, one or more easily excited electrons. Such atoms give rise to spatially isolated energy levels within the otherwise forbidden band gap. An electron in such a state can be thermally excited into the conduction band by the acquisition of much less energy than that required to excite it directly from the valence band. These donor states, as they are called, donate electrons to the conduction band, and thereby increase the conductivity of the material. Materials that show electrical conductivity largely as a product of such states in the forbidden band gap are termed impurity semiconductors.

An impurity semiconductor may also be an acceptor type. To explain this term it is desirable to first introduce the concept of a "hole." In Fig. 4, the transition labeled I is that of an electron in the valence band being excited directly into the conduction band, where it can contribute to the electrical conductivity of the crystal. Note, however, that this migration leaves an unoccupied state in the valence band (labeled p in Fig. 4). This absence may be considered as a bubble or hole in a sea of electrons. At any particular moment this hole is localized spatially. However, it may wander through the crystal because an adjacent electron may jump into this vacant energy state to leave behind a hole where it had been; then, another electron can drop into this new vacancy to leave a new vacant state behind it, and so on. If an external electric field is impressed upon the material, a hole will tend to move in the opposite direction of the electron drift, that is, toward the negative electrode, thus conducting electricity. This process is called hole conductivity.

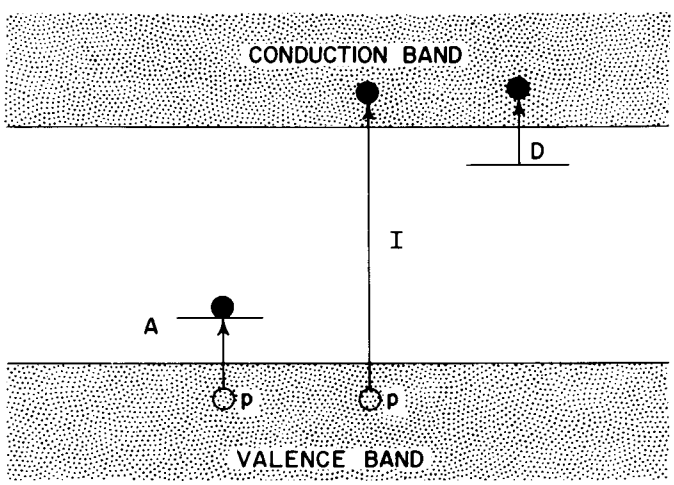


Figure 4. Impurity Semiconductor

There are subtleties here which cannot easily be explained without recourse to the mathematics of quantum mechanics, but for the purposes of this article it is adequate merely to note that a hole is an absence of an electron in a normally filled band, and that it behaves very much like a particle of roughly electron mass and electron charge, but with opposite sign, i.e., positive instead of negative.

The transition labeled D in Fig. 4 is that of an electron from a donor state into the conduction band. If an impurity has fewer electrons than are necessary to satisfy bonding requirements in the lattice into which it is introduced, then it is electron hungry and can act as an acceptor of electrons. Such impurities usually produce energy levels within the forbidden band gap that are fairly near the valence band energy-wise. Electrons, therefore, can be thermally excited from the valence band into these states to leave behind a hole which can contribute to the conductivity of the material by hole conduction. This process is shown in Fig. 4 as the transition labeled A. A crystal having its electrical conductivity due chiefly to conduction electrons is termed N-type; if its electrical conductance is due chiefly to hold conduction, it is termed P-type.

The reader may wonder why these impurity atoms give rise to discrete states in the forbidden band gap rather than bands. The reason hinges on the amount of impurity. If it is assumed that the impurity atoms are fairly uniformly distributed, they will be several lattice spacings apart and therefore, the coupling between them will be weak. Because it is this coupling between atoms which gives rise to the bands, weak coupling will cause insignificant splitting of the impurity levels. Thermal vibrations of the lattice can broaden these impurity levels, as can high concentrations of impurities, but these effects will be ignored in this presentation.

Thermal excitation is not the only way electrons can be excited from the valence band to the conduction band. Light, if its energy is greater than that of the forbidden band gap, can do the same. The energy of a light quantum is given by $h\nu$, where h is Planck's constant, equal to 6.62×10^{-27} erg-seconds and ν is the frequency corresponding to two wavelengths of the light. The relation between wavelength and frequency is $\lambda = c/\nu$, where c is equal to 2.998×10^{10} centimeters per second (the velocity of light) and λ is the wavelength in centimeters. A very useful relation to remember is the following equation relating the wavelength and the energy of a light quantum given in electron volts:

$$(\lambda \text{, in microns})\ (E \text{, in ev}) = 1.24 \qquad (1)$$

It is apparent that the maximum wavelength which will excite an electron from the valence band to the conduction band is given by:

$$\lambda \text{ max (in microns)} = 1.24/\text{Band gap (electron volts)} \qquad (2)$$

As an example, for a band gap of 2.0 electron volts, the impinging light must have a wavelength shorter than 0.62 micron (μ) (6200 Angstroms) for the material to manifest photoconductivity. Because of certain complicating factors, photoconductivity may be initiated at somewhat longer wavelengths, but this possibility need not be of concern for the moment.

If the wavelength of a quantum of incident light is short enough to excite an electron to the conduction band, its movement will leave a hole in the valence band. As long as the electron and hole are free, they can contribute to the conductivity of the material. However, these free electrons and free holes may, if they get close enough to each other, recombine, i.e., the electron may drop into a vacant state in the valence band. Due to the relatively small density of carriers, as well as considerations arising from the principle of conservation of momentum, such direct recombination is an exceedingly improbable occurrence. However, in all crystals localized energy states exist in the forbidden band gap introduced by imperfections. These imperfections may be vacant lattice sites; interstitial atoms (i.e., atoms within the crystal, but not at lattice sites), impurity atoms in either substitutional or interstitial positions; dislocations, that is, mechanical imperfections in the crystal; or one of several other types. In any case, such imperfections give rise to discrete energy states in the otherwise forbidden band gap. These states are localized — not only in terms of energy, but spatially — within the crystal.

If an electron in the conduction band, while wandering through the crystal, approaches close to one of these localized energy states, it may lose a small amount of energy by exciting lattice vibrations, and drop into this localized energy state — become trapped by it. If this state, now filled by an electron, then captures a free hole, recombination will have taken place, for a free electron and a free hole will have annihilated each other. Such a localized energy state is said to be acting as a recombination center. If these discrete states are close to the conduction band, however, there is an appreciable chance that a trapped electron may be thermally re-excited into the conduction band before that state traps a hole, in other words, before recombination takes place. The previously trapped electron is then again available for conduction purposes and has not been lost completely as in the case of recombination. Such a state is acting as an electron trap.

Analogously, a level near the valence band which traps a hole may have this hole thermally re-excited into the valence band before an electron is trapped. Such a level is then acting as a simple hole trap rather than as a recombination center.

Fig. 5 shows both shallow trapping states which act as hole traps or electron traps, and those deeper-lying states which act predominantly as recombination centers. It should be remembered that probabilities are being dealt with here. The probability of thermal excitation varies exponentially with energic distance from the allowed bands and will change from a high probability to a low probability in a narrow energy range, roughly kT wide. Although it is not strictly correct, we may take what is shown in Fig. 5 as the steady-state Fermi levels to be demarcation levels between those states acting predominantly as simple trapping states and those states acting predominantly as recombination centers. These

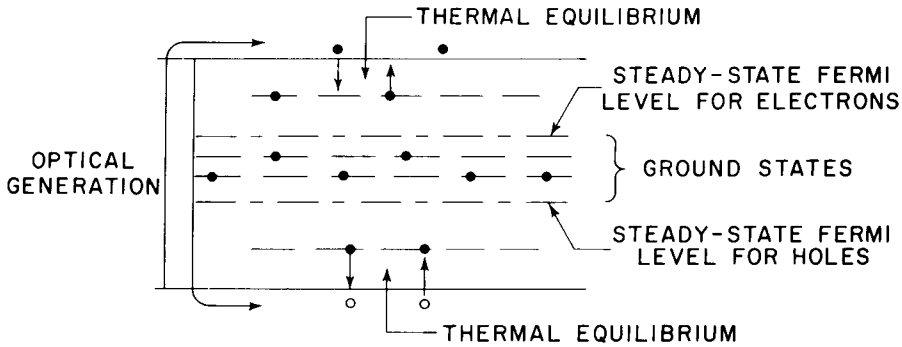


Figure 5. Model Showing Both Shallow Trapping States and Ground States

latter, situated deep within the forbidden band gap, are termed ground states.

It is apparent that thermal excitation and photoexcitation across the band gap are inverse processes to recombination. Excitation puts electrons into the conduction band, and holes into the valence band. Recombination takes them out. In the steady state, the rate of recombination must equal the rate of excitation. The rate of recombination varies directly with the number of unoccupied ground states and the number of conduction electrons. Other things being equal, the number of conduction electrons will vary inversely with the number of unoccupied ground states. The significance of this relationship will appear later, when applications of the theory are discussed.

APPLICATIONS

Based on physical state, photoconductors can be divided into three large groups. These are: (1) single-crystal photoconductors, (2) microcrystalline photoconductors, and (3) amorphous photoconductors.

Much has been published concerning single-crystal photoconductors; models have been constructed, theories have been expounded. A good deal less has been written about microcrystalline photoconductors, and very little is available on amorphous photoconductors, which are the ones of greatest importance in vidicon applications. There are very practical reasons for this lack. A single crystal without imperfections is the most amenable to theoretical analysis. A crystal without imperfections is an impossibility, so even if the "perfect crystal" problem could be solved completely, the range of its applicability would be restricted. Any deviations from the ideal crystal — any imperfections introduced — increase the analytical and computational difficulties until only with very gross approximations and questionable assumptions can the problems be solved.

In passing from the single-crystal state through the microcrystalline state to the amorphous state, the density of imperfections in the crystal lattice increases by several orders of magnitude. Is it any wonder that theorists cling to the single crystal? From a practical point of view, the tube designer can not so limit himself because some of the most important photoconductive devices, such as sintered photocells and the vidicon, are based on these less-perfect states. However, in the application of the simplified theory previously given, the order of complex to more complex will be followed. The single-crystal germanium and cadmium sulfide photoconductors will be discussed first.

Germanium

Single crystals of germanium used in photocells are grown from the melt and zone-refined in much the same fashion as crystals grown for transistor use. The germanium is doped with arsenic, which is in group V of the periodic table (see Table I) and thus has one more valence electron than it needs for bonding requirements in a germanium (group IV) lattice. Arsenic, then, introduces donor-type energy states near the conduction band of the germanium. The extra electron of the arsenic is easily lost into the conduction band and the germanium becomes more conductive. Arsenic-doped germanium is, thus, N-type because the majority of the carriers are electrons.

Table I

Periodic Table							
Groups	I_b	II_b	III_b	IV_b	V_b	VI_b	VII_b
			B	C	N	O	F
			Al	Si	P	S	Cl
	Cu	Zn	Ga	Ge	As	Se	Br
	Ag	Cd	In	Sn	Sb	Te	I
	Au	Hg	Tl	Pb	Bi	Po	At

In one application, thin slices (about 20 mils thick) are cut from a single-crystal germanium ingot and then subdivided into small rectangles 160 by 95 mils. These pieces will be the photosensitive elements, but first, a good deal of processing is required. The previous mechanical treatment has work-hardened the surface and therefore greatly increased the number of surface recombination states. If it is recalled that in the steady state (other things being equal) the number of conduction electrons varies as the number of unfilled recombination centers, then it is reasonable that a large number of surface recombination states will reduce the photoconductivity. Because such reduction is undesirable, about one half of the 20-mil thickness is electrolytically etched away. The etching removes the work-hardened

portion and considerably improves the surface condition.

An indium dot is placed on the center of the germanium wafer which is then baked at 550 C for 8 to 10 minutes in a hydrogen atmosphere to diffuse some of the indium into the germanium. Indium is in group III of the periodic table and thus is shy one electron for germanium-type bonding. It introduces acceptor levels into the forbidden band gap of the germanium. These levels lie near the valence band. Electrons in the valence band can easily be thermally excited into the indium acceptor levels to leave a hole which is free to wander about in the valence band. Of course, some of these states obtain their electrons from the arsenic donor states and thus the effect of the arsenic is neutralized. Going from the germanium toward the indium metal, the conductivity passes from N-type to intrinsic, where the indium has just neutralized the effect of the arsenic, to P-type and, upon reaching the pure indium, to metallic. The indium dot serves as one electrode; the other electrode is attached to one edge of the wafer, often by tin-soldering (see Fig. 6).

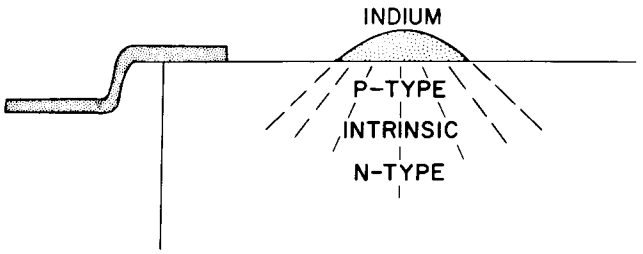


Figure 6. P-N Junction Photocell

The result of this processing is a P-N junction type of photoconductor. Its operation is illustrated by Fig. 7 which depicts an N-type material with donor levels near the conduction band joined to a P-type material with acceptor levels near the valence band. In thermal equilibrium, electrons which are concentrated in the conduction band of the N-material tend to diffuse into the conduction band of the P-type material where their concentration is low. Similarly, the holes in the valence band of the P-type material will diffuse into the valence band of the N-type material, where holes are scarce. However, whereas the negative charges of the conduction electrons were compensated by the positive charges of the donor atoms in the N-type material, there is no similar charge compensation in the P-type material. Therefore, electrons diffusing into the P-type material will make it negative. Similarly, whereas the positive holes in the P-type material were compensated by the negatively charged acceptor ions, they will, when they diffuse into the N-type material where there is no such charge compensation, make it more positive. The diffusion of electrons and holes from where they are concentrated to where they are not sets up a dipole layer, positive on the N-side and negative on the P-side,—a condition which prevents any further net diffusion of carriers. Note that in Fig. 7 the Fermi levels in both materials are aligned. The Fermi energy is related to the chemical potential of the materials. It can be shown from considerations of thermodynamics that for thermal and electrical equilibrium to exist, the Fermi levels of all materials in contact must be equal.

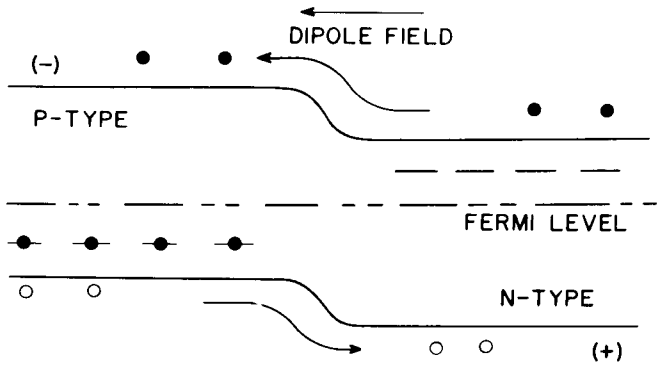


Figure 7. P-N Junction Band Diagram

If an external field, such as to make the P-type material more negative, is applied, the height of this potential barrier will be increased by an amount proportional to the external field. Due to the high barrier, the diffusion current will be quite small, just enough to balance the current due to thermally excited carriers. Because electrons are scarce on the P-side and the electrons can't get over the barrier, the current due to the impressed field will also be very small. A similar conclusion can be drawn for holes. Then the effective resistance in the dark will be high. However, if an electron-hole pair is generated by light on the P-side and the electron is within a diffusion length of the barrier, it can plummet down the barrier into the conduction band of the N-type material and move toward the positive electrode; the hole will move toward the negative electrode. Similarly, if an electron-hole pair is generated on the N-side and the hole is within a diffusion length of the barrier, it will float up into the valence band of the P-material toward the negative electrode, while its electron moves toward the positive electrode. The sum of these currents constitutes the photocurrent.

The germanium cell is a P-N junction and must be operated with the P-side negative, that is, with reverse voltage. Were it to be operated in the forward direction, the dark current would be too high for the device to be useful.

Cadmium Sulfide

Unlike the germanium cells, the cadmium-sulfide, single-crystal photocells are not P-N junctions, but rely on a wide band gap (2.3 ev compared with germanium's 0.75 ev) to provide a low dark current. One method of processing this type of cell is to grow the crystals of cadmium sulfide in a vapor-phase reaction between cadmium vapor and hydrogen sulfide. The crystals are formed at a temperature of 960 to 980 C and immediately after formation are doped with a halide by moving the crystals into a temperature zone of 600 to 700 C and then passing nitrogen containing one of the hydrohalide acids over them.

The resistivity of the undoped cadmium sulfide crystal might be about 10^{12} ohm-centimeters. After the material is doped with halide, the resistivity of the crystal drops to about 1 ohm-centimeter. Along with this

resistance change, the sensitivity (measured in terms of the number of charges passed through the crystal per photon absorbed) is increased. In order to increase the resistivity, the crystals are cooled, soaked in a very dilute solution of a copper salt such as cupric chloride, and then washed, dried, and baked in a static hydrogen-sulfide atmosphere at 700 C for 5 or 10 minutes to enable the copper to diffuse into the crystal. This treatment increases the resistivity of the material and decreases the decay time, i.e., the time required for the signal to decay the dark-current level after the exciting light is turned off; it also decreases the sensitivity somewhat.

The variations described above can be explained in terms of the theory. A pure crystal has high resistance, low sensitivity, and short lag or decay time. The resistance is high in the dark because the band gap of cadmium sulfide is large (2.3 ev) and very few electrons acquire enough thermal energy to overcome it. Because of the general imperfections in the crystal, a fairly large number of recombination centers exist which keep the number of conduction electrons low. This scarcity reduces the sensitivity and, in addition, causes the decay to be fast, because the moment the light is cut off, the recombination centers quickly gobble up the conduction electrons, which are no longer being replenished by photoexcitation.

The addition of monovalent halide ions, which substitute for the divalent sulfide ions, introduces donor levels near the conduction band because the halide has one more electron than it needs to satisfy its bonding requirements in the lattice. It can be predicted from theory that not only the halides, but also elements from group III of the periodic table will do this. Such atoms as aluminum, gallium, indium, or thallium, which could substitute for the cadmium, all have one more electron than is necessary for bonding with the sulfide ion and would, therefore, introduce donor levels.

The donor levels, introduced into the forbidden band gap either by halide or group III elements, lie sufficiently close to the conduction band that almost all of them are ionized at room temperature. The conductivity is, therefore, enhanced. There is another effect, though: namely, that some of the available electrons fill up recombination centers in the dark. Thus, when the crystal is illuminated, fewer unfilled recombination centers are available and the photoresponse is therefore, increased. This effect, plus the fact that the empty donor sites can act to some extent as traps which must be filled upon illumination and emptied afterwards — processes which take time — causes the rise and decay times to be increased.

When silver or monovalent copper is added, it substitutes for the cadmium. The silver atoms having one less electron than necessary for bonding, introduce acceptor states near the valence band. These states take up the excess electrons of the previously introduced halide or group III element and thus neutralize its effect very much as the indium acceptor states did when introduced into germanium-containing arsenic donor states. The resistivity then rises, the lag is decreased, and the photosensitivity is somewhat decreased. The exact amounts of the two doping agents used are determined empirically and vary somewhat with the use to be made of the photocell.

Microcrystalline Photoconductors

Microcrystalline photoconductors are represented by the powder photoconductors, generally used with a plastic binder, and by sintered photoconductors. The latter can be prepared, for example, by spraying an aqueous slurry of a finely ground photoconductor onto a suitable substrate, drying the surface, and then firing it to form a polycrystalline sintered layer. Doping materials are much the same as for single-crystal photoconductors, but they are generally added to the slurry before spraying rather than introduced after crystal growth.

The microcrystalline photoconductors present applications of the simplified theory similar to those of the single crystals and will not be discussed further.

Amorphous Photoconductors

The most important current commercial use of amorphous photoconductors is as the light-sensitive element in vidicon-type television camera tubes. The remainder of this article discusses the application of amorphous photoconductors to these tubes.

Although many chemists have seriously questioned the existance of an amorphous state, (i.e., one in which there is no order) because of their belief that there always exists at least some short-range order, even in liquids, this classification will be retained here. Eschewing a theoretical definition, it will be said that a material is amorphous if it does not give an X-ray diffraction pattern containing lines. Amorphous photoconductor targets for the vidicon are presently prepared by evaporation in a poor or good vacuum, depending on the type of surface desired. To understand the characteristics required of a photoconductor for use in a vidicon, it is important to understand the principles of operation of this type of tube. Fig. 8 illustrates the operation of a vidicon.

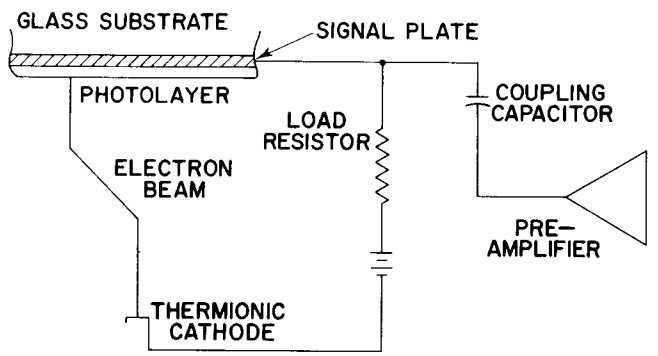


Figure 8. Operation of Vidicon

The glass faceplate of the vidicon is coated with a transparent, conducting layer. This material might be tin oxide (TIC, NESA, as it is variously called), indium oxide, or one of the other conducting oxides,

or a thin evaporated layer of some metal, as for example, gold. Such considerations as spectral transmission, conductivity, chemical reactivity with the photoconductor which is laid down upon it, propensity for diffusion into the photolayer, electronic characteristics of its contact with the photoconductor, durability, and ease of preparation affect the selection of the particular conducting layer used. The photoconductor substrate presently in vogue is tin oxide (TIC), although the use of bismuth oxide-gold layers is increasing. The photoconductor, which in the case of current commercial vidicons is antimony trisulfide, is evaporated onto this substrate. In operation, the photolayer is scanned by a low-velocity electron beam. The beam side of the photolayer becomes stabilized at essentially cathode voltage. Since the conducting substrate, which acts as a signal electrode, is maintained, for example, at a voltage of 50 volts above that of the cathode, the assembly may be considered a form of sandwich cell, with the transparent conducting layer as one electrode and the electron beam as the other.

If a light image falls on the photosurface, the conductivity of the photoconductor will be modified; it will be greatest in the areas where the intensity of illumination is greatest. During the interval between coincident scans (1/60 second for broadcast TV pictures) charges transferred transversely across the photolayer will cause the potential to rise on the beam side. The portion of the beam electrons landing on the photolayer on a successive scan is determined by the resulting potential pattern. The signal results from capacitive coupling to a load resistor of the charges landing on the beam side of the photosurface coupling is obtained through the signal electrode. The signal is taken from the load resistor, amplified, and used. More information on vidicon operation can be found in the article entitled "Camera Tube Design and Processing." (Also, see Refs. 1 and 2).

PHOTOCONDUCTORS FOR THE VIDICON

The basic performance characteristics of the vidicon are determined in large part by the characteristics of its photoconductive target. The properties of photoconductive surfaces, pertinent to their application in vidicons are:

1. Spectral response
   a. Peak wavelength
   b. Half-width
2. Transfer characteristics
   a. Sensitivity
   b. Gamma
3. Rise and decay characteristics
   a. Rise time
   b. Lag, stick
   c. After-image, burn
4. Dark current-voltage characteristics
   a. Operating voltage
   b. Signal-to-dark current ratio
   c. "k" value
5. Stability
   a. Chemical
   b. Physical

Spectral response describes the variation of signal with wavelength for a given illumination or, alternatively, the illumination required to obtain a given signal at any wavelength. The peak-wavelength and half-width terms are self-explanatory.

The vidicon can be thought of as a transducer which changes illumination information into electrical information. Its transfer characteristics characterize the relationship between the illumination on the photoconductor and the electrical signal obtained from it. The most fruitful method of presenting the transfer characteristics is on a log-log plot — the logarithm of the signal vs. the logarithm of the illumination. A straight line on such a plot represents a power law variation, that is a relationship in which the logarithm of the output signal is proportional to the logarithm of the illumination. The slope of such a straight line is known as gamma, a photographic term indicative of contrast.

Rise and decay characteristics describe the time required for a change in the illumination on the photosurface to manifest itself in a corresponding change in the output signal. The rise time is the time required for the signal to rise to within 1/e (where e is the base of natural logarithms) of its equilibrium value after the light level is changed (usually from zero, i.e., the dark condition). This definition implies an exponential rise which is not always found. Lag and stick are arbitrary measures of the decay time and apply to short-term effects (about a second or less). It is found empirically that, for the usual television scan rates, the picture will smear when the image is moved over the photosurface if the signal does not decay to less than 15 per cent of its initial level within three scans (3/60 second). * After-image and burn refer to long-term effects. The after-image is measured by the time required for an image to become imperceptible after all incident illumination is removed from the photosurface. Its usual range is from two seconds to several minutes. Although usage varies, burn is generally restricted to those cases in which an appreciable signal remains after a quarter of an hour or so and which can be removed only by stringent measures such as washing it out with uniform illumination of high intensity, operating the tube with high dark current, or increasing the target voltage until secondary emission stabilizes the surface potential near that of the mesh.

Another important photosurface characteristic is the relation between the dark current and the target voltage. As in the case of the transfer characteristics, it is convenient to plot this characteristic on log-log paper. The logarithm of the dark current is plotted against the logarithm of the target voltage, i.e., the voltage applied to the signal electrode. Over an appreciable

*Note: By a "scan" is meant a single pass of the beam of a given point. Because of the relatively large diameter of the beam in the usual vidicon, there is no effective interlacing, and a scan occurs every 1/60 second.

portion of its range, such a plot will approximate a straight line. The slope of this line is termed the k value. For antimony trisulfide surfaces, the customary range of k is from 2.0 to 4.0, although both higher and lower values are found. If a surface has a k value much above 3.0 its dark-current variation is apt to be excessive; slight variations in the thickness of the photolayer cause variations in the electric field across the photoconductor which, in turn, when k is high, may give rise to appreciable variations of the dark current over the surface.

The stability of a photoconductor is of prime importance in any application. A photoconductor which is chemically unstable and prone to change its electronic characteristics on life is obviously undesirable. The red form of selenium, for example, is metastable and must be operated below 40 degrees Centigrade if it is to be kept from rapidly converting to the thermodynamically more stable gray form. Indium tri-iodide is deliquescent and must be kept in a dry atmosphere during processing. Antimony trisulfide surfaces will not tolerate the customary tube bakeout at a temperature of 400 C but revert to the black, conducting form at a temperature in the neighborhood of 200 C. Some of the low-melting photoconductors, which appear interesting because of their relatively fast decay times, cannot be used successfully because their vapor pressures are so high at the operating temperature of the vidicon that they evaporate rapidly from the photosurface in the vacuum of the tube.

Most of the properties of the photoconductive target listed at the beginning of this section are determined by the choice of photoconductors. Nevertheless, radical variations in these properties can be produced by modifications of the chemical preparation of the basic material and by co-evaporation with other materials, as well as by suitable choice of evaporation parameters. Since antimony trisulfide (Sb_2S_3) is presently most in use as the basic photoconductor for vidicons, it will be discussed in detail. Other photoconductors will be considered later.

Antimony Trisulfide Photosurfaces

There are, basically, four kinds of antimony trisulfide photosurfaces. These are:

1. Compact (frequently termed "solid")
   Evaporated in high vacuum (10^{-4} millimeters of mercury or less)
2. Porous
   Film density less than about 0.5 gram per cubic centimeter
3. Quasi-Porous
   Film density in the range from about 0.5 grams per cubic centimeter to 2.0 grams per cubic centimeter
4. Composite
   Includes combinations of the above three, as well as co-evaporation with other materials

The general characteristics of a compact, (high-vacuum evaporated) antimony trisulfide surface are: high sensitivity, high lag (long decay), high capacitance, low blue sensitivity, high dark current, and a tendency to give secondary emission at relatively low target voltages. A porous layer, on the other hand, is characterized by lower sensitivity, lower lag, low capacitance, improved blue response, relatively low dark current, and a low secondary-emission ratio. The quasi-porous layer partakes variously of the characteristics of both compact and porous layers, depending on the evaporation pressure as well as certain other evaporation parameters. The most common composite layer is a compact layer evaporated on a porous substrate. This combination increases somewhat the dark current and lag over those of a porous layer, but the main effect is to enhance the sensitivity.

The high capacitance of the compact layer (an average layer has a capacitance of over 12 millimicrofarads per square centimeter) is due in part to its thinness (roughly one to two microns) and, in part, to its high dielectric constant (over nine). As may be seen from Fig. 9B, the density of a porous layer is only a few tenths of a gram per cubic centimeter compared with over four grams per cubic centimeter for a compact layer. Thus, in a porous layer, the effective dielectric constant is much reduced and the capacitance of a typical layer is only about one millimicrofarad per square centimeter. That portion of the lag due to the capacitance of the target and the effective beam resistance (the so-called "capacitance" lag) is, therefore, appreciably reduced in the porous layer. Actually, except in very thin compact layers or barrier-type layers, the photoconductance lag predominates over the capacitance lag.

Effects of Pressure

Fig. 9A depicts the variation of the electric field required for a standard dark-current density through the photolayer with the ambient gas pressure during evaporation. Note in particular the negative slope of the curve above about a pressure of 150 microns and the sharp drop below that pressure. The quasi-layer is obtained by evaporation in the transition region between compact and porous. While the general shape is maintained, the curve may be moved up or down and the position of the transition region shifted slightly by changes in photosurface thickness, evaporant flux density, evaporation geometry (primarily throw distance), substrate temperature, composition of starting material, and nature of the ambient gas. In Figs. 9A and 9B, note the coincidence of the sharp drop in field required to sustain a standard dark-current density and the equally sharp increase in the density of the photolayer.

Fig. 10 shows the variation of the dark-current density with target voltage for representative compact, porous, quasi-porous, and ASOS photosurfaces. The last-named is a special composite layer which will be discussed later. Representative transfer characteristics for these four kinds of photosurfaces are given in Fig. 11.

The characteristics of the deposited layer have been

shown to be affected by the nature and pressure of the residual gas or any vapors from contaminants in the system. Because pressures of the order of 10^{-10} millimeters of mercury are required to reduce the rate of residual gas-substrate collisions well below the rate of evaporant-substrate collisions, even at high evaporation rates, really gas-free photoconductor layers have not been investigated. Nevertheless, it seems desirable to reduce the residual gas pressure to as low a level as possible in the evaporation of solid layers and, in particular, to avoid metallic and organic contaminants. Thus, the standard preferred high-vacuum practices should be adhered to, including such things as the avoidance of cadmium or zinc-plated parts, the use of welded or silver-soldered joints rather than lead-soldered joints, and the use of an oil-diffusion pump and a liquid-nitrogen trap.

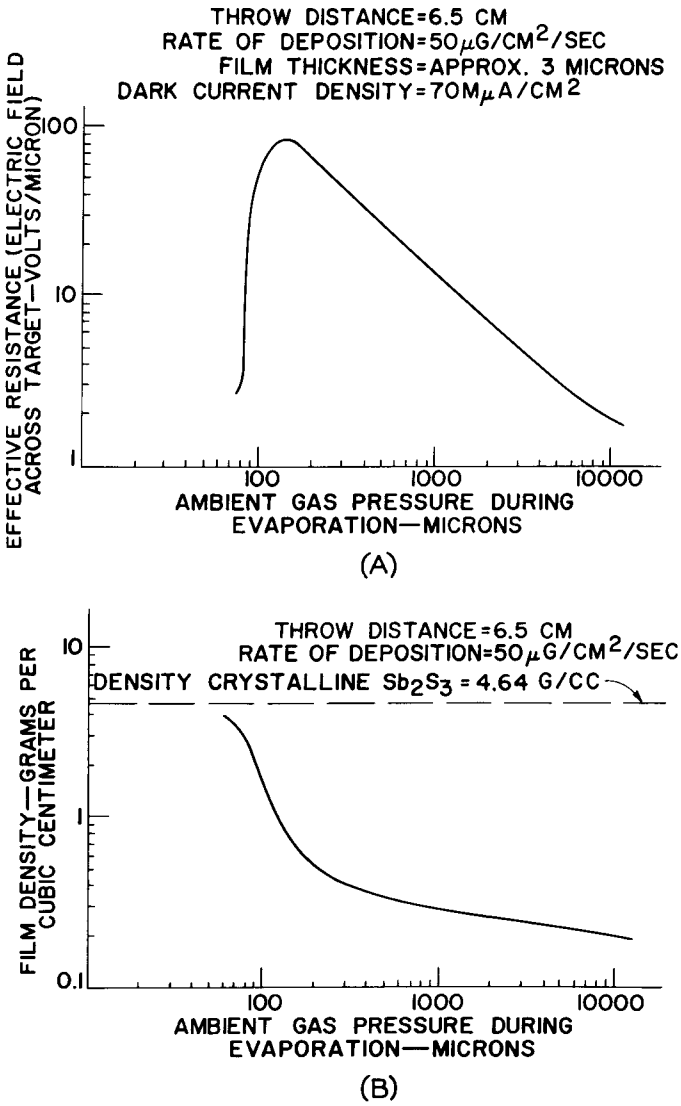


Figure 9. Effects of Pressure: (A) Effective resistance vs. gas pressure during evaporation (B) Film density vs. gas pressure during evaporation

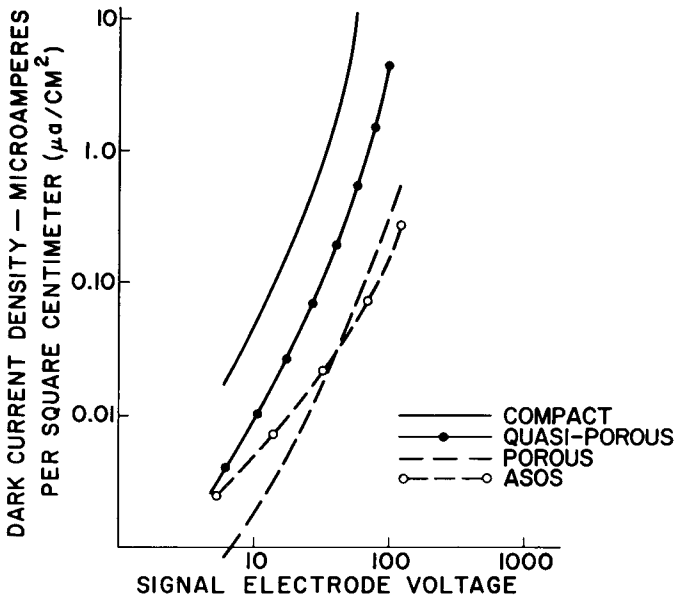


Figure 10. Plot of Dark Current Density vs. Signal Electrode Voltage

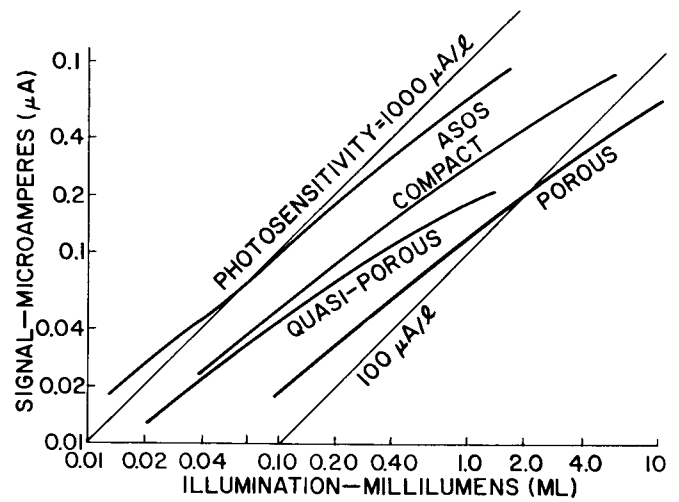


Figure 11. Signal-Illumination Transfer Characteristics

Effects of Photolayer Thickness

Fig. 12 shows the spectral response for two different thicknesses of high-vacuum-evaporated antimony trisulfide. Note that the spectral response peak shifts from 0.5 micron in a thin layer to over 0.65 micron in a thick layer. This red shift is due to the absorption characteristics of the antimony trisulfide which, since it absorbs strongly in the blue region, allows only red light to affect the deeper layers. Because the blue light can penetrate more easily into the porous layer by taking advantage of the interstices, even relatively thick layers of the fluffy porous Sb_2S_3 do not show a very large self-absorption red shift. Thus, the response peak of the porous layer lies near that of a very thin compact layer.

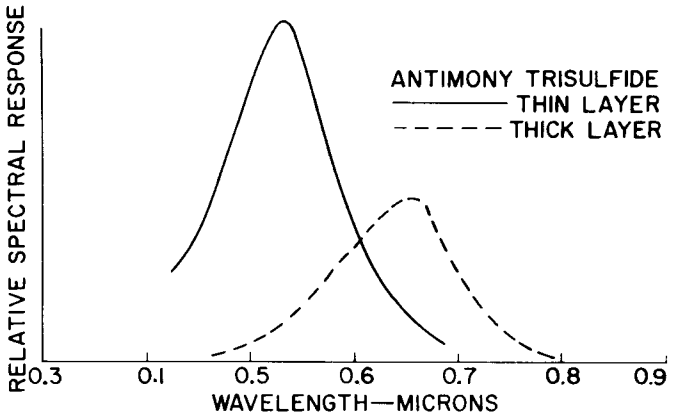


Figure 12. Variation of Spectral Response with Film Thickness

When the thickness of the photoconductor is greater than the average distance that the carriers travel before recombination, the sensitivity is reduced. As the thickness of the photolayer is increased, a larger proportion of the carriers generated by the illumination is swallowed up in the interior of the material without being able to reach an electrode and contribute to the conductivity.

Part of the observed lag in the vidicon is due to the capacitive effect of the target. A thick layer generally has a slight advantage in lag over that of a thin layer, because the former has a lower capacitance.

Effects of the Substrate

Some of the requirements imposed on the substrate have been previously mentioned. Not yet mentioned, but easily the most important requirement is that the photoconductor adheres firmly to the substrate. Porous layers generally present no problems but there are severe limitations for compact layers, especially those which must be exposed to the atmosphere at some time during tube processing. The mechanical bond of the photoconductor to the substrate is determined by many factors. One of the most important is substrate temperature. Fig. 13 illustrates film stress for a typical semiconductor film evaporated on a glass substrate. Each line represents a constant substrate temperature.

It is apparent that only a small range of temperature exists which permits thick layers to be evaporated onto the substrate without peeling of the layer. What this temperature range is depends on both the substrate and the photoconductor. As yet, no quantitative work has been done on antimony trisulfide evaporated on TIC, but what evidence there is indicates that the optimum temperature lies below the temperature at which the antimony trisulfide layer converts to a black, highly conducting form, and above normal room temperature.

During the course of the evaporation process the substrate temperature tends to rise. If a substrate heater is used and a constant substrate temperature is to be maintained, the power input to the substrate must be reduced as the evaporation progresses. Recognition of this fact is crucial in the preparation of thick layers.

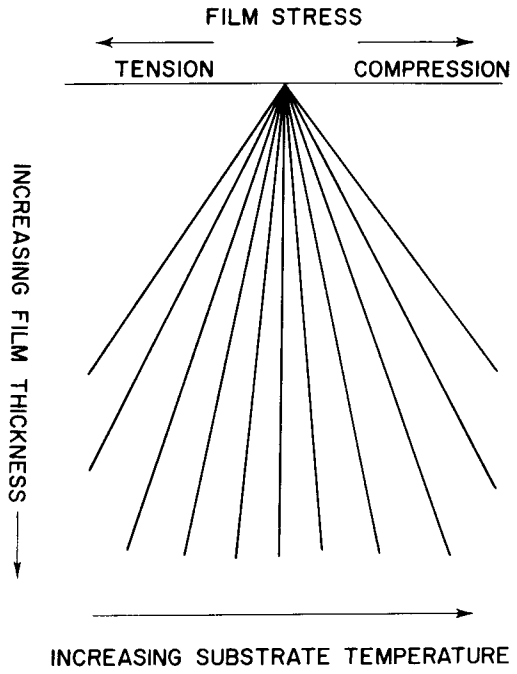


Figure 13. Film Stress for a Typical Semiconductor Evaporated on Glass Substrate

Change of substrate temperature also changes the accommodation coefficient of the surfaces exposed to the evaporant stream. It is well known that the amount of material reflected from the interior surfaces of the evaporation chamber to the target is appreciably increased if the temperature of these surfaces is raised, even though they are below the evaporation temperature of the evaporand. It has also been found that an increase of substrate temperature tends to make the resulting photolayer more dense and shifts the porous-compact transition range toward higher pressures. If there is more than one constituent in the evaporand, the composition of the resulting photolayer will be a function of the substrate temperature: even if only a single component is present, phase changes may be caused by an elevated temperature or by surface quenching, i.e., by the sudden cooling of the evaporant as it deposits on the substrate.

Scrupulous cleanliness of the substrate is a requisite for the adherence of thick photoconductor layers. Great care must be taken to avoid even the least residuum of soap or detergent. Organic solvents which tend to leave films (acetone is a frequent culprit) should either not be used, or else the substrate should be cleaned after their use. If the final rinse is with water, distilled water should be used. However, to avoid water spots, it is usual to follow the distilled-water rinse with a vapor rinse. Although a steam jet can be used for this purpose, one of the most satisfactory procedures is to place the clean, still-wet target blank in a saturated atmosphere of an organic solvent miscible with water. Acetone or an alcohol meets this requirement and is usually used. The reservoir of solvent is maintained somewhat above room temperature. The solvent vapor condenses on the target blank, rinses it and removes residual water in the process. When thermal equilibrium is reached the target blank will be essentially dry

and may be removed. If evaporated gold is to be used as the signal electrode, a very clean substrate may be obtained by cleaning the glass target by standard glass-cleaning procedures, followed by distilled water and vapor rinses. Then, before the gold signal electrode is deposited by evaporation in a vacuum, the glass is subjected to a glow discharge, which is one of the best methods of cleaning glass surfaces. Unfortunately, it also destroys most of the transparent conducting coatings now in use and, therefore, cannot be used to advantage after the signal electrode is applied.

The quality of the substrate is important in the prevention of spots and blotches which show up in the vidicon picture. These defects can be caused by granularities, pinholes, high-resistance areas, or surface irregularities in the signal electrode. Water spots also show up in the picture, as do insulating particles lying on the surface.

Evidence suggests that, in addition to affecting adherence and spots, substrate character and quality can have a considerable effect on the burn characteristics of a photoconductive surface, especially on negative image burn which is occasionally encountered. The choice of the substrate and its methods of subsequent processing are critical factors in the over-all performance of the photosurface and should not be slighted.

Effects of Modifications of The Basic Photoconductor

It has been known for many years that variation from stoichiometry of the Sb_2S_3 used in the preparation of vidicon targets affects the characteristics of the resulting photosurface. It was early found that a slight excess of antimony in the starting material (greater than the 71.8 per cent in stoichiometric Sb_2S_3) improves the over-all performance of the porous and the porous-compact kind of composite layer. While photosurfaces evaporated from material containing an excess of sulfur, may have fair sensitivity (largely a result of the very high target voltage required to obtain the standard dark-current density), their decay characteristics will be poor. Image retention and burn are particularly severe, even in the stoichiometric material. As the amount of antimony in the material is increased, the sensitivity rises and the decay characteristics improve. At some percentage, apparently somewhat above the Sb_2S_3-Sb eutectic (73.1 per cent Sb), the over-all performance becomes optimal. The lag continues to decrease slightly as the excess of antimony is further increased, but the sensitivity drops and the long-term decay (image retention and burn) is increased. Roughly, the same variations occur in the case of compact Sb_2S_3 photosurfaces, but here additional factors enter, for, as the proportion of antimony is increased over the stoichiometric percentage, the tendency for thick layers to peel is accentuated. In addition, the conversion to the dark, highly conducting form of Sb_2S_3 takes place at a lower temperature. Furthermore, while the reduction of target voltage (for standard dark-current density) with increased antimony content is generally not a problem in porous layers, it is a severe limitation for the compact layers. Solid layers have a low target voltage (show high conductivity) and further reduction sharply reduces the maximum signal obtainable (a function of the target voltage, capacitance, and scan rate) and the sensitivity, and makes the target more prone to show the effects of beam-landing errors (porthole effect).

Another method of modifying the basic photoconductor is to intentionally introduce impurities into the material during preparation but reliable data have not been obtained on the effects of such doping. Most experiments indicated either no effect or a slight worsening of the over-all performance; in a few experiments some improvement was seen. Doping with silver, monovalent copper, and arsenic have at various times appeared to improve the characteristics of the photolayer, but (until serious effort is expended in trying to optimize amounts and methods of introduction) great improvement does not appear likely.

Miscellaneous Evaporation Parameters

The geometry of the evaporation system has a marked effect on the evaporation process. The throw distance, i.e., the distance from the evaporator to the target, directly affects the uniformity of the deposit. The expected evaporant distribution pattern on a stationary target from a point source can be easily derived for a high-vacuum evaporator. However, the design of most evaporators is such as to make them more or less directive and, even if the dimensions of the evaporator are small compared with the throw distance, the effective size of the source may be considerably enlarged by the collision of evaporant molecules with each other and with residual gas molecules. It has been computed that, for a moderately fast rate of evaporation, the local effective pressure due to evaporant molecules is greater than 10 microns.

In a high-vacuum evaporation system, such factors as the accommodation coefficient of the target substrate and the walls of the evaporation chamber are important. To reduce the reflection of evaporant molecules from the walls of the evaporation chamber onto the target, the chamber itself should have as large dimensions as convenient and should be kept as cool as possible. Sand blasting the exposed surfaces of the evaporation chamber increases the effective accommodation coefficient and thus reduces reflections and discourages peeling which contribute to spots on the photoconductor. It has the disadvantage of making cleaning of the chamber more difficult.

Generally, if the throw distance is kept large compared to the dimensions of the target, and nonuniformities of substrate temperature are avoided, it is not too difficult to obtain moderately uniform high-vacuum layers. Obtaining uniformly porous surfaces is much more difficult. Although this important process has not been completely investigated, some observations can be made concerning it. Examination of a porous photosurface with an electron microscope discloses that the layer is composed of relatively large aggregates of roughly spherical particles, 100 to 200 angstroms in diameter. These aggregates, in turn, form chains and larger aggregates which can be seen under 1000X magnification. The surface of the porous photolayer is highly structured and has discernible buttes,

plains, and valleys. The variation in elevation may, in very porous layers, exceed one-third of the total layer thickness. It may be that the ability of the electron beam to penetrate some distance into this structure accounts for the decrease of electric field required for a given dark-current density with decrease in the density (increase of porosity) of the photolayer. This explanation is further suggested by Fig. 14, which shows that the drop in the field with decreased density is most pronounced in thin layers where the fraction of the total thickness that the beam could be expected to penetrate is greatest.

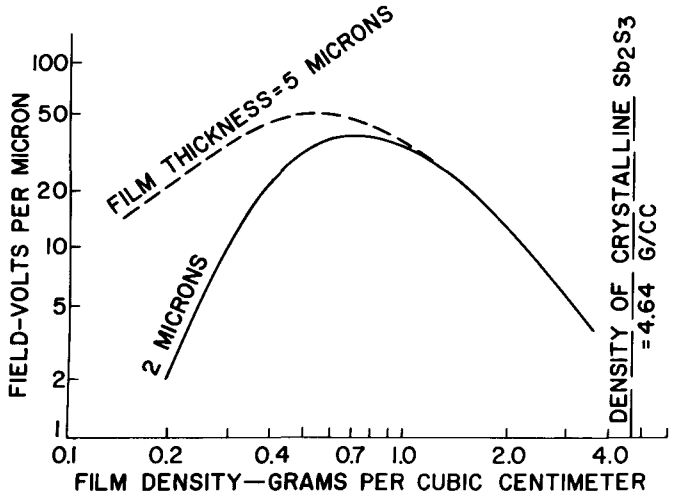


Figure 14. Electric Field Required to Support a Dark Current of 0.07 Microamperes per Cubic Centimeter Plotted Against Density of the Photolayer

It is known that the formation of the small "building blocks" (100 to 200 angstroms in diameter) disclosed by the electron microscope is necessary if a porous photolayer is to be deposited. What part the further agglomeration of these particles in transit plays in the formation of the porous layer is not clear, although light-scattering experiments indicate that particles whose diameter is at least a few tenths of a micron do exist in the evaporant cloud.

Evaporant aggregates in a poor vacuum behave somewhat similarly to smoke. Above an ambient pressure of a few hundred microns, evaporation chamber geometry and temperature gradients induce convection currents; these can be observed by the use of fine beams of light. Much turbulence is in evidence and, in some configurations, vortices are established which can give rise to annular deposits of photoconductor on the target. At lower pressures, the influence of convection currents is not as obvious. The course of travel of the evaporant particles may be more influenced by the mean free path of the aggregates than by their velocity per se. Shield temperature plays an important role in determining the distribution of particles in the evaporant stream.

It follows that, within the confines of a small tube such as the 1-inch vidicon, it is extremely difficult to obtain uniform porous layers. A way of bypassing this difficulty is to perform the evaporation on the target (generally the faceplate of the vidicon) in a chamber of more favorable geometry and then to cold-seal the faceplate to the bulb. While a large bell jar could be used to achieve layer uniformity, its use would result in large amounts of photoconductor being wasted and in the photoconductor evaporant permeating the entire vacuum system and making cleaning of the system difficult. A preferred method is to employ a relatively small evaporation chamber to achieve a degree of lateral uniformity by the use of a long evaporator boat and longitudinal uniformity by maintaining a constant evaporation rate and causing the faceplate to travel across the top of the evaporation chamber in a line normal to the axis of the boat. This principle is employed in the so-called traveling-plate evaporator.

Use of the traveling-plate evaporator offers other advantages in addition to achieving uniform layer thickness. The photolayer is more homogeneous because, during the traverse of the target, there is little time for the nature of the evaporation to change. The evaporation parameters can be controlled more closely and the evaporation system is easier to clean. Monitoring a glass plate which travels ahead of the target allows close control of the thickness of the layer and permits evaporator temperature or traverse speed to be adjusted before the target passes over the evaporation chamber. By sacrificing monitoring time, it is feasible to form several targets in succession.

The rate of evaporation is important in deciding the porosity of the evaporated layer (see Fig. 15). Note that rate of deposition, rather than rate of evaporation is used as abscissa; the rate of evaporation is not directly measurable. Both rates will be functions of ambient pressure and evaporator-boat temperature. The rate of deposition will also be a function of the throw distance. The data for Fig. 15 were taken with pressure constant and throw distance fixed.

The sharp rise in the density of the evaporated layer at very low rates of deposition is evidence that there must be a sufficient number of evaporant-evaporant collisions for nucleation to take place before a porous layer can be formed. This statement is true regardless of the fact that, due to collisions between the evaporant and the ambient gas, the energy with which the evaporant strikes the substrate may be reduced to a relatively low value. A glance at Fig. 9B will reveal that a sufficient number of evaporant-ambient gas collisions must also have occurred for the evaporant energy to be reduced to the level at which nucleation will take place. At any given rate of evaporation above the minimum necessary for nucleation, there is a narrow range of ambient pressure at which the transition from a compact layer to a porous evaporated layer will take place. As might be expected, this range is also affected by the nature of the ambient gas. Helium, with its low molecular weight and long mean free path, will, at a given pressure, permit a denser layer to be deposited than, for example, nitrogen or argon.

Because the evaporation rate is a direct function of the evaporator-boat temperature, control of evaporator-boat temperature is vital in achieving reproducible photosurfaces. Over a large portion of the ris-

ing part of the curve of density vs. rate of deposition (Fig. 15A), the density varies roughly as the square root of the rate of deposition. While this is not a rapid variation, it requires only a small change of boat temperature to produce quite a large change in rate of evaporation. By evaporating at rates corresponding to the minimum of this curve, it is feasible to reduce the influence of boat temperature on the density of the resulting photosurface.

The throw distance (evaporator-target distance) strongly affects the porosity of the evaporated layer. In evaporations at several hundred microns pressure, it is possible to effectively terminate deposition by increasing the throw distance a few inches. It is apparent that to produce a layer of a given density, as the throw distance is decreased, it is necessary to decrease the pressure. Fig. 15B illustrates the strong dependence of film density on throw distance. Note the progression of the curves with decreasing ambient gas pressure. As the gas pressure is reduced the inflection point of the density-throw curve moves toward a greater throw distance.

A lower limit to the throw distance is imposed by temperature effects on the substrate and the deposited photoconductor, as well as by considerations of photosurface uniformity. An upper limit is imposed by practical size of the evaporation chamber. In addition, at any particular pressure, the rate of deposition becomes sufficiently low at large throws to effectively limit the throw distance. Fig. 15C shows this effect strikingly, because at increased pressures the dependence of the rate of deposition on throw distance becomes very strong. While it is likely that optimum throw distances exist for various values of the other evaporation parameters, this subject has not been completely investigated.

The evaporator itself is an important factor in the evaporation process. Fig. 16 illustrates several evaporator styles. The first two, 16A and 16B, are cups wound with molybdenum or tungsten wire. Because most hard glasses release contaminants at the temperatures used for evaporating many photoconductors, Vycor or fused quarts is preferable for use as the cups. From the point of view of chemical reactivity, such evaporators are excellent. However, both their thermal inertia and the energy input required are large. In the simple design shown, the hot heater wires are exposed to the back diffusion stream of evaporant and this exposure may cause some dissociation. In addition, the high energy input heats up the evaporation chamber and in some cases gives rise to reevaporation of material deposited on the walls. The target is also exposed to this excess heat and while heat treatment during and after evaporation may in some cases improve the characteristics of the photosurface, such heat treatment is best effected in a more controllable fashion.

Fig. 16C shows an evaporator cup made with alumina-covered wire; this type has proved satisfactory in some applications. Its thermal inertia is high, but lower than that of the Vycor evaporator. The type shown in 16D can be made very small, however, spattering of the

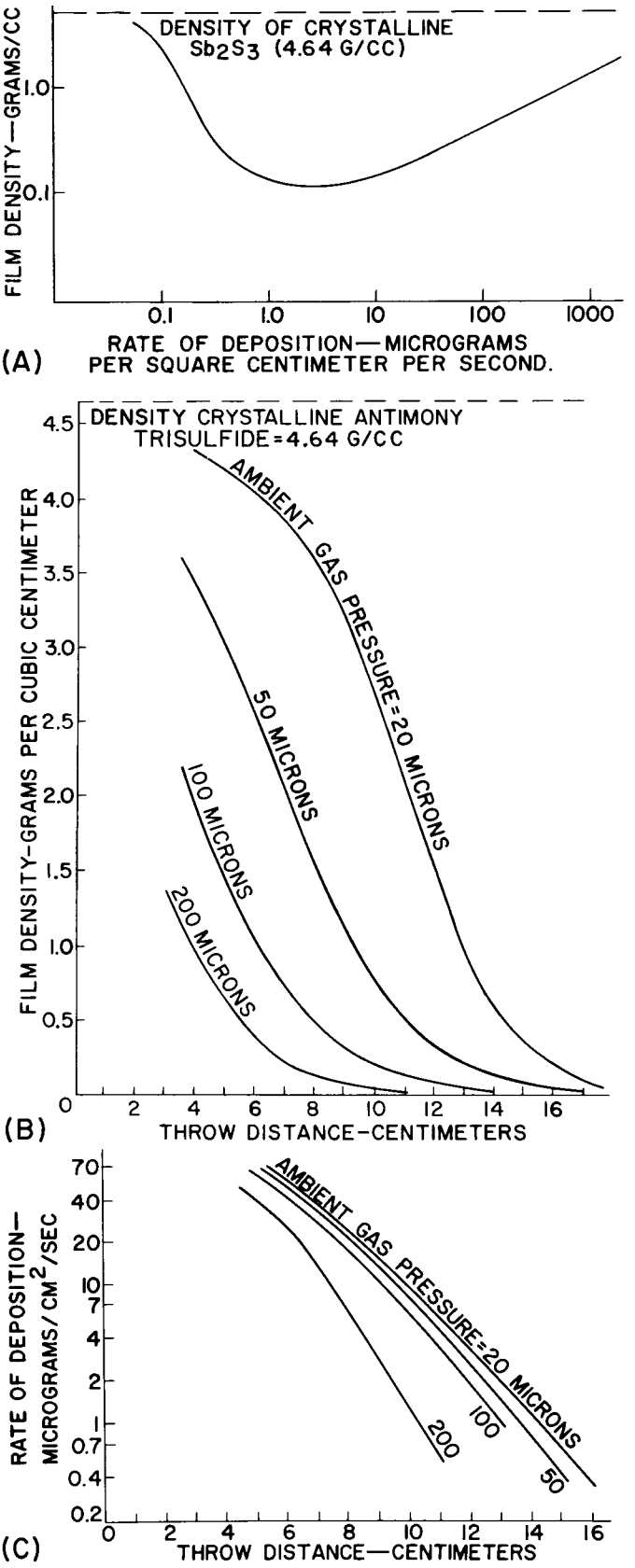


Figure 15. Variation of Film Density and Rate of Deposition: (A) Film density vs. rate of deposition, (B) Film density vs. throw distance with ambient evaporation gas pressure as a parameter, (C) Rate of deposition vs. throw distance

photoconductor is apt to occur because of the depth of the cup. The directly heated tantalum boats illustrated in Fig. 16E, F, and G have very low thermal inertia. Their power consumption is small compared with that for indirectly heated cups and satisfactory evaporations can therefore be made at a lower temperature. For most materials, this type of boat is more or less satisfactory from the standpoint of chemical reactivity. Evaporation of a more inert metal into the boat can improve this characteristic if thermal diffusion or evaporation is not excessive. The cross section of the directly heated boats should be adjusted to obtain uniform temperature at the operating point. Thermal conductivity down the leads, radiation losses, and cooling due to the evaporation of the photoconductor are factors which need to be considered. A particular advantage of the boat illustrated in Fig. 16G is that the substrate is not in direct line with the material and is, in fact, exposed only to the vapor. This type of evaporator is of considerable aid in reducing the incidence of spots on the photosurface caused by evaporant spatter.

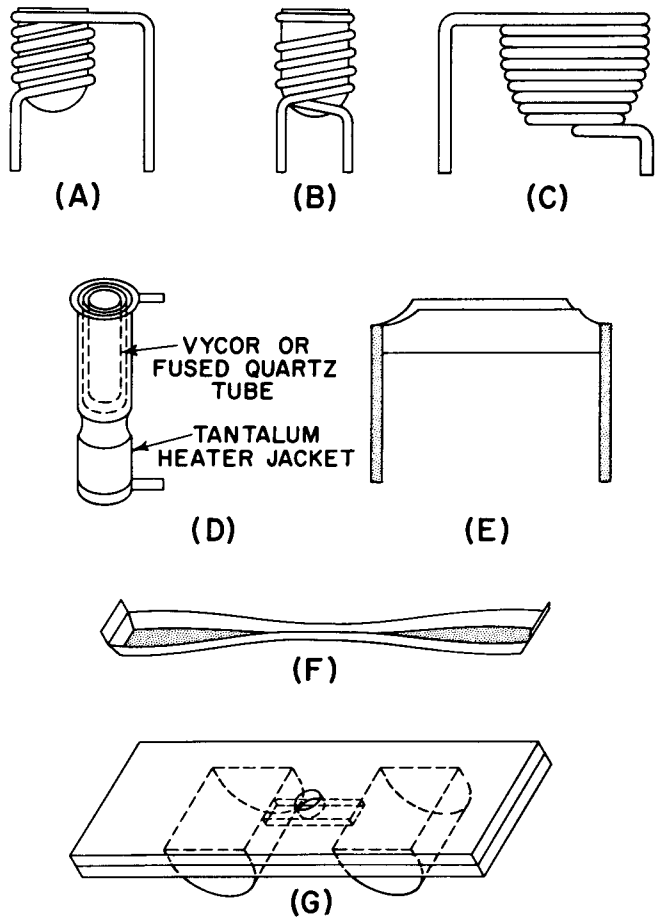


Figure 16. Types of Evaporator Boats: (A) Single-helix-wound Vycor cup, (B) Double-helix-wound Vycor cup, (C) Alumina-coated molybdenum or tungsten wire, (D) Tantalum-sheath Vycor evaporator, (E) Temperature-compensated tantalum boat, (F) Linear tantalum boat for traveling-plate evaporator (temperature compensated), (G) Anti-spatter point-source tantalum boat for high-vacuum evaporation

FACTORS INFLUENCING SELECTION OF PHOTOCONDUCTORS FOR THE VIDICON

General Requirements of Photoconductors for the Vidicon

The choice of photoconductors for use in the single-layer vidicon photosurface is limited by the desired spectral response of the tube. This subject will be discussed in detail later in this article. In addition, the scan rate, target area and thickness, dielectric constant, and resistivity are interrelated in such a way as to further restrict choice. For example, if T be the time between scans, R the resistance, and C the capacitance of the photolayer, then for frame storage, the following inequality must hold.

$$RC >> T \tag{3}$$

If V is the potential rise on the beam side of the photoconductor, then the highlight signal current which can be obtained is

$$I = CV/T \tag{4}$$

This signal must be greater than a certain minimum I_{min} in order that a sufficiently high signal-to-noise ratio be obtained. In the vidicon, I_{min} is determined largely by the noise generated in the preamplifier stages and the quality of the picture desired.

It would appear that both C and V should be as great as possible. However, both are limited by other conditions. The potential rise V is determined by the relation

$$V = V_T(1 - e^{-T/RC}) \tag{5}$$

where V_T is the target (signal electrode) voltage.

The resistance R is a function of the amount of light on the photosurface. The particular functional relationship is characteristic of the particular photolayer.

Fig. 17 shows the potential rise V in the dark (0) and at three different light levels (1, 2, and 3). After each scan time T, the beam returns the potential of the beam side of the photosurface essentially to cathode potential (taken here as zero volts). The difference between the potential for a light level (1, 2, or 3) and that for the dark level (0) is a measure of the resulting signal. It is apparent that for highlights, the signal approaches saturation, that is, a very small increase in signal results from even a large increase in light. It is further apparent that if the dark resistivity of the photolayer were so low that during one scan time the potential V rose to V_T (for example, if the dark-current curve was at 2 or 3, then the maximum light signal obtainable would be quite small for a moderate value of V_T. It is this consideration that determines the relation given in Eq. (3).

The obvious palliative, increasing V_T, is useful only to a degree, for when V_T is greater than, say 100 volts, secondary-emission and beam-bending effects frequent-

ly degrade the picture. Also, because it is the charge deposited by the beam that reduces the potential of the photosurface, the useful potential rise V is limited by the ability of the beam to discharge the target.

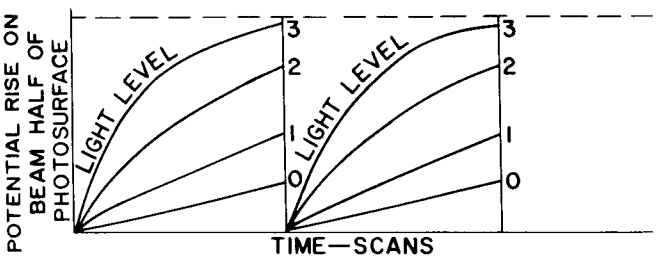


Figure 17. Potential of Beam Side of Photo Surface vs. Time

The capacitance of the target is limited by requirements of short decay time. If the illumination on the photosurface is removed, the potential on the beam side, instead of rising after scanning to its illuminated equilibrium value, will be scanned down until it oscillates between its minimum value and that corresponding to its dark equilibrium value. Because the velocity distribution of the beam causes it to have, in effect, a high resistance, it will in general require more than one scan for the potential on the beam side to reach equilibrium, even if the "photoconductive" lag were zero. The smaller the capacitance of the target, the smaller will be this "capacitance" lag effect.

If end effects are ignored, the capacitance C of a vidicon photolayer is given by the equation

$$C = kA/4\ \pi\ d \quad \text{(cgs units)} \qquad (6)$$

where k = dielectric constant of the layer, A = the area of the scanned surface, and d = the thickness of the layer

Thus, the target capacitance can be reduced by decreasing the raster area (e.g., by going from a 1-inch to 0.5-inch vidicon), increasing the photolayer thickness, or by reducing the effective dielectric constant (by going from a compact layer to a porous layer).

By the use of Eq. (6), Eq. (3) can be expressed in terms of more-basic properties of the photolayer. The resistance of the layer is given by

$$R = \rho d/A \qquad (7)$$

where ρ = the resistivity of the film

Thus, in practical units,

$$\left(\frac{\rho d}{A}\right)\left(\frac{kA}{4\pi d}\right)\frac{1}{8.99 \times 10^{11}} = 10^{-13} k\rho \gg T \qquad (8)$$

As an example, for compact Sb_2S_3, k = 10. Thus, $\rho \gg 10^{12}/60$ ohm-meter or, very approximately, $\rho > 10^{12}$ ohm-centimeter.

An intrinsic photoconductor with this resistivity would have a band gap such that the threshold for light absorption would be in the far red. As will be seen later, such a photoconductor would have its peak sensitivity in the visible spectral region. It is fortunate that the peak occurs in the visible region. If the peak occurred elsewhere, the requirements of spectral response might preclude successful operation of the simple single-layer vidicon and force the use of targets having much more complex structures.

Spectral Photoconductive Response

The theory discussed earlier suggests a relationship between the band gap, absorption spectrum, and spectral photoconductive response of a material; such a relationship exists although it is not as precise as one would like.

In a metal, light of very long wavelengths (i.e., of low energy) can excite electrons for the same reason that thermal excitation is easy: there are empty energy states immediately above the filled ones. Metals, then, are strongly absorbing not only in the visible region, but also in the infrared region.

On the other hand, a pure insulator requires light of very high energy to excite electrons from the valence band to the conduction band and thus absorbs only in the ultraviolet region and transmits visible radiation. However, in the region in which the insulator does absorb, it is photoconducting, because the additional electrons excited from the valence band to the conduction band will contribute to the conductivity of the material. There is, however, a group of compounds whose conductivity is not increased by the absorption of light. These are the compounds of atoms with only partially filled "d" shells. Light shining on such compounds as iron oxide, nickel oxide, or the manganese oxides, for example, can raise electrons to a higher energy state, but the electrons and holes either remain bound to each other, forming a so-called "exciton," or have extremely short lifetimes and so do not contribute to the conductivity. In the pure state, these compounds have not been observed to be photoconductive, although they absorb in the visible part of the spectrum and as a result are highly colored.

Disregarding such compounds, there is a rough method of estimating the spectral response of a photoconductor. As a rule of thumb, it can be said that "Photoconductivity begins in earnest at that wavelength at which the material begins to absorb radiant energy strongly." Fig. 18 illustrates this rule. Note that the inflection points of the transmission spectra of antimony trisulfide and indium trisulfide lie very near the peak of their photoconductivity curves. Note, however, how poorly this rule of thumb applies to selenium. Fortunately, few photoconductors are as far off as selenium. However, this example does show that the rule should be applied with caution.

The question may arise as to why the photoconductivity drops at shorter wavelengths where the absorption is almost 100 per cent. The explanation is similar to that used to explain the drop in sensitivity of thick

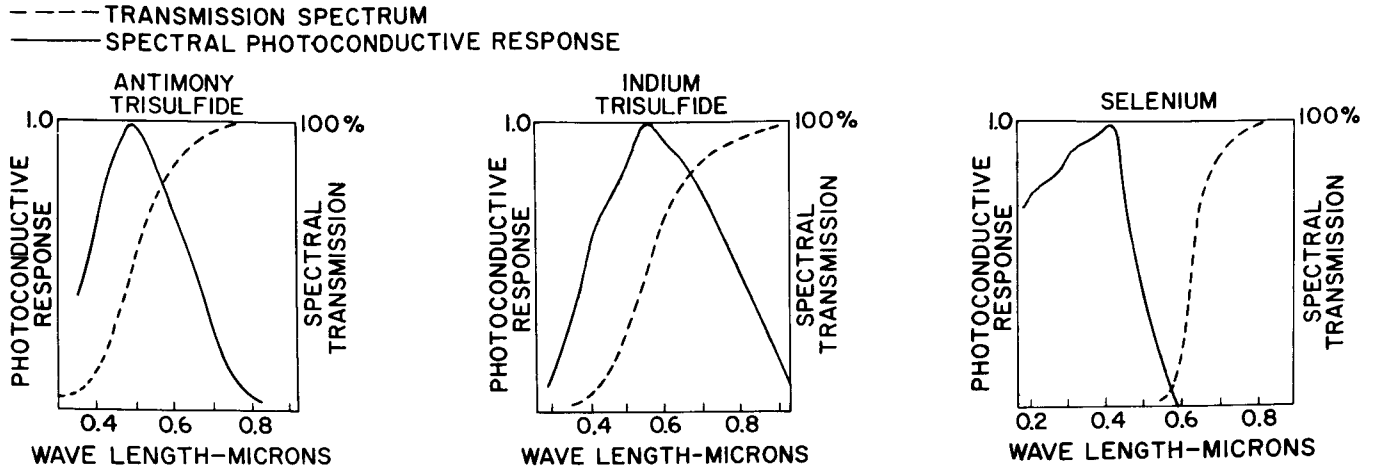


Figure 18. Comparison of Transmission Spectrum with Photoconductive Response for Three Photoconductors

layers of antimony trisulfide. At short wavelengths, most of the light is absorbed in a thin layer near the surface of the photoconductor. Here, the electrons and holes have a high density and a correspondingly greater rate of recombination. Furthermore, any carrier which goes to the opposite electrode must make its way through a maze of obstacles such as traps and recombination centers for almost the entire interelectrode distance. If, as happens frequently, the range of the carriers is short compared to this distance, very few will get to the opposite electrode unless they are generated some distance below the surface, i. e., unless the light penetrates well into the material. Thus, if the light is absorbed very strongly, the photocurrent will drop, as it is seen to do in Fig. 18.

To apply the rule-of-thumb relationship between photoconductive spectral response and special transmission, the spectral transmission must be known. It can be obtained roughly from considerations of color, if the incomplete "d" shell exceptions are remembered. If a material is white or colorless, it transmits in the visible region. If photoconductivity in the visible region is desired, such transmission is not desired, but the material might make an ultraviolet photoconductor. If the material is black, it probably absorbs so strongly in the visible region as to give negligible photoconductivity in this region. However, such material may be a good infrared photoconductor, especially if it is cooled below room temperature to reduce the dark current due to thermal excitations.

Based on color, many of the elements and compounds in the I, II, III, IV, V, VI, and VII "B" groups of the periodic table appear as possible photoconductors. See Table I and Fig. 19. Regularities and trends in such properties as ionicity and homopolarity, heats of formation, and interatomic distances of binary compounds throughout the periodic table are well-known to chemists. Similar trends can be found in the transmission spectra, which may be used as a more accurate parameter than simple color.

The wavelengths (in microns) of the inflection points of the spectral transmission curves of a series of binary compounds of cadmium, gallium, and indium are plotted in Fig. 19; the binary compound is indicated on the abscissa. From left to right, following the C's the plot shows cadmium fluoride, cadmium chloride, cadmium bromide, cadmium iodide, then cadmium oxide, cadmium sulfide, cadmium selenide, and cadmium telluride. Similar treatment is used for gallium (denoted by S) and for indium (denoted by X).

From Fig. 19, it is possible to estimate the color of the compound. Those compounds lying below the lower limit of the visible region are white or colorless. Those lying above the upper limit are black, approaching metallic. Those lying between the limits, but near the lower limit are yellow by transmitted light, because they absorb only in the blue region, i. e., they absorb all wavelengths below their ordinate on the chart. Those higher in the visible region appear red by transmitted light because they absorb both blue and green.

The fluorides and chlorides of cadmium, gallium, and indium could be ultraviolet photoconductors at best. Their iodides, oxides, sulfides, and selenides might be good photoconductors for visible radiation. Indium iodide, for example, is nearly panchromatic, as is its sulfide. The tellurides of these compounds, on the other hand, although infrared photoconductors, absorb so strongly in the visible region that their sensitivity is low there. Note than gallium oxide is missing from the chart. By extrapolation, the inflection point of the transmission curve for gallium, and hence, by the rule of thumb, its photoconductive response peak, would lie in the ultraviolet region. The compound itself is found to be white, which reinforces the validity of the extrapolation.

Regularities are to be found in many other series of compounds. Their importance in this discussion lies in their usefulness for extrapolation and interpolation where there are no data available for the compound in question. The trends in the spectral transmissions are reflected very well in corresponding trends in the spectral photoresponse. Fig. 20 shows that the spectral response peaks of cadmium sulfide, cadmium selenide, and cadmium telluride occur exactly as would

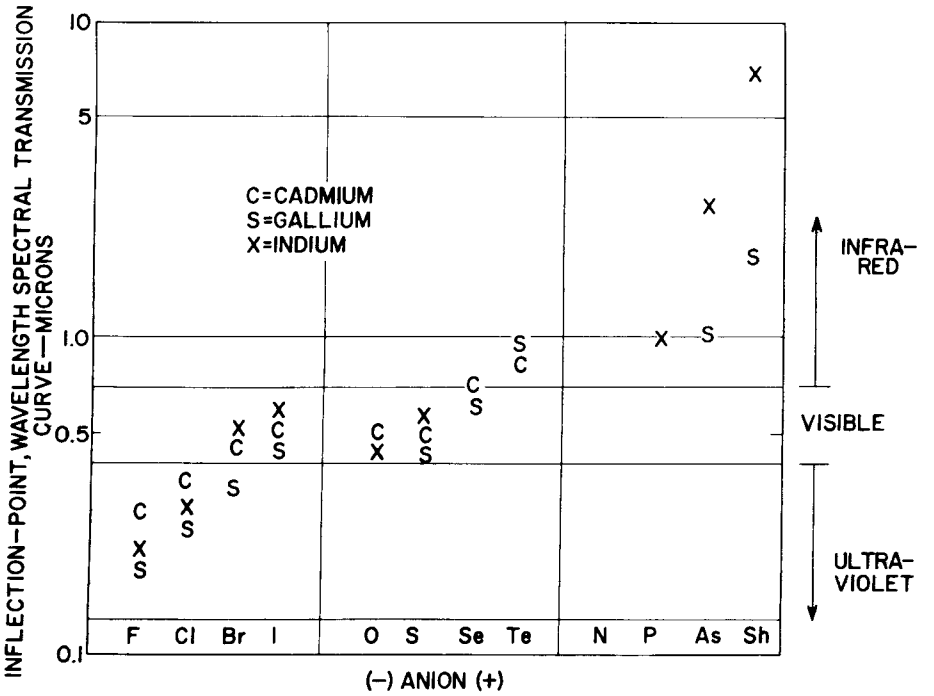


Figure 19. Regularities in Binary Compounds of Cadmium, Gallium, and Indium

be expected from an analysis of Fig. 19. Cadmium sulfide peaks at about 0.5 micron or 5000 angstroms, cadmium selenide at 0.72 micron, and cadmium telluride in the near infrared region at 0.82 micron. The trend also works in another way. In Figs. 19 and 20 the cation was kept constant and the anion changed. If the anion (the more electro-negative element) is kept constant and the cation is changed, a trend will also be in evidence. Zinc and mercury are in the same group (II B) as cadmium. The trends in their physical and chemical properties are reflected in a red shift in the photoconductive response as the series progresses from zinc to cadmium to mercury. Note in Fig. 21 that the spectral response peak of zinc sulfide is about 0.3 micron whereas that of cadmium sulfide is 0.5 micron. Zinc sulfide is white; cadmium sulfide, yellow; and mercury sulfide, red. The peak for mercury sulfide is even further into the red region than cadmium sulfide; in fact, it is close to where the zinc telluride curve lies. Mercury selenide is black and its response peak is in the infrared region.

There are bound to be exceptions to simple rules for complex materials. Fig. 22 illustrates one. The curves for the lead selenide and telluride are just the reverse of what they would be expected to be. Most of the other properties, for example, the optical dielectric constants and major absorption peaks in the intrinsic absorption spectra do follow the expected order. All three curves (for sulfide, selenide, and telluride) lie in the infrared region above 2 microns and are quite close together. The anomaly in the photoconductive response can probably be attributed to the complex structure of the bands. While most photoconductors exhibit a negative-temperature dependence of band gap having a value between -3×10^{-4} and -5×10^{-4} electron volts per degree, the lead group exhibits an almost equal, but positive dependence. In other words, as the temperature of most photoconductors is increased the band gap becomes smaller and the response peak is shifted toward the red. For the lead group (sulfide, selenide, etc.), the band gaps become larger as the temperature is increased and their response peaks are shifted toward the blue.

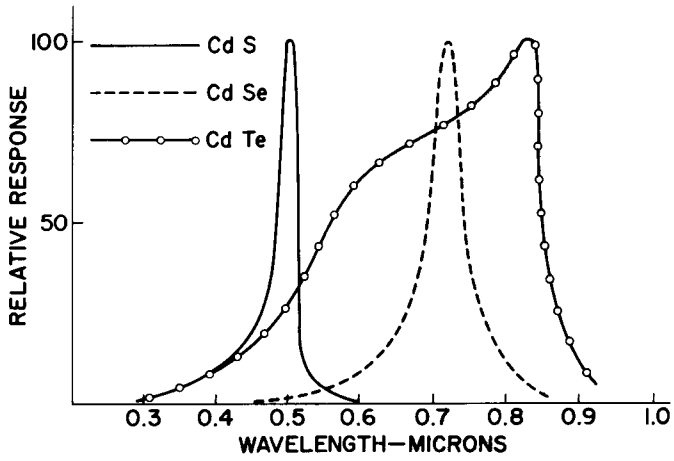


Figure 20. Spectral Response of Cadmium Sulfide, Selenide, and Telluride

Excluding the rare exceptions, such as the lead group, selenium, and the unfilled "d" shell group, the transmission spectrum or color is a fairly reliable means of selecting photoconductors on the basis of where it is that their photoconductivity is wanted. To a considerable extent, this test will roughly determine the dark conductivity of the intrinsic material because, in general, the greater the band gap, the greater the dark resistivity. Doping and crystal imperfections, however, can radically alter this relationship as in the case of cadmium sulfide single crystals doped with copper and chloride where the resistivity can change by over 12 orders of magnitude.

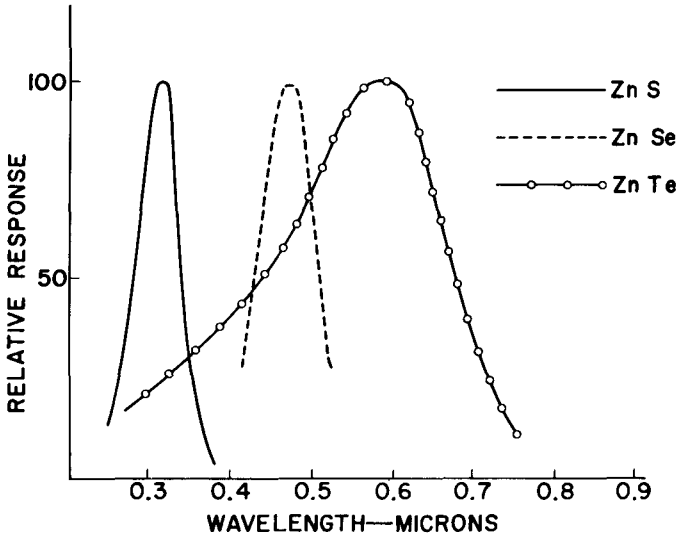


Figure 21. Spectral Response of Zinc Sulfide, Selenide, and Telluride

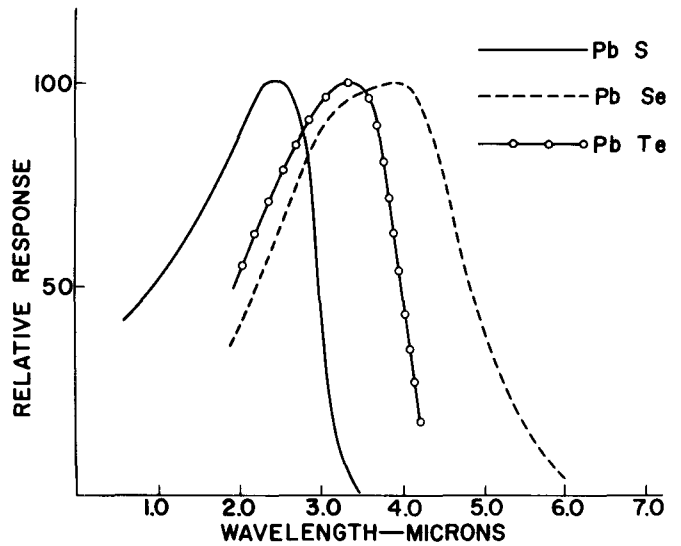


Figure 22. Spectral Response of Lead Sulfide, Selenide, and Telluride

The melting points and boiling points of the material (matters of very practical importance) show the same sort of trends as do the inflection points of the transmission curves. Unfortunately, two of the most important aspects of the photoconductivity of a material — the sensitivity and the decay — cannot be estimated directly from a consideration of the constituent elements either by the simple theory used here or by any other known to date. One further interesting and sometimes useful correlation is that existing between heats of formation and forbidden band gaps. The correlation is not as good as that between spectral transmission and band gap, but heats of formation can frequently be found where transmission spectra are lacking.

COMPOSITION GRADED PHOTOSURFACES

Mention was made earlier of pre-doping the material to be evaporated. It is also possible to treat the finished photosurface by exposing it in an atmosphere of an active agent or by baking it in such an atmosphere. Air bakes and H_2S bakes have been tried with Sb_2S_3 and a few other photoconductors with varying success. A third method of doping the photosurface is to co-evaporate the doping agent and the photoconductor. Co-evaporation may also be used to construct graded junctions of two or more photoconductors.

A graded-junction photosurface is made by controlling the deposition rates of two or more materials throughout the progress of the evaporation. The process produces a composition gradient across the thickness of the photolayer. The electronic characteristics of such composition-graded photosurfaces are frequently quite different from those of any of the constituents. An example is the so-called ASOS layer which, although it is a compact (high-vacuum evaporated) layer consisting largely of Sb_2S_3 may be more resistive than a porous Sb_2S_3 layer and more sensitive than a compact Sb_2S_3 layer. Depending on its composition gradient, an ASOS photolayer may have a decay significantly faster or slower than a solid Sb_2S_3 layer. Furthermore, this decay may vary considerably depending on the wavelength of light to which the photolayer is exposed; the decay is longer for shorter wavelengths. By choosing materials whose evaporated layers are distinctly P or N type, it might be feasible to construct PN, NP, NPN, and PNP graded junctions by controlled co-evaporations.

The degree of control required to successfully reproduce composition-graded photolayers is extremely high. Slight variations in the deposition rates are sufficient to markedly modify the electronic characteristics of the resulting photolayer. The price for improved performance is greatly improved control of the evaporation parameters.

A device which has proved of considerable value in monitoring composition-graded photolayers is the so-called "Pin-Hole Evaporation Camera" (PHEC). It is useful only for high-vacuum evaporations (below 10^{-5} millimeters of mercury), where the mean free path of the evaporant molecules is appreciably greater than the dimensions of the evaporation chamber. Fig. 23 shows the device schematically.

During the evaporation, "images" of the evaporators are deposited on the monitor plate. The thickness of these images builds up in proportion to the amount of each material deposited on the faceplate. The image thickness may be monitored by transmitted light or, through the use of interference bands, by reflected light. A typical plot of interference bands vs. evaporation time is shown in Fig. 24.

By taking the slope of each curve at a particular time and by adding the ordinates, the relative rate of evaporation of each component at a particular fraction

of the total photolayer thickness can be obtained. By doing this for several points, a plot of composition vs. fractional thickness can be made. Fig. 25 shows such a plot for the evaporations illustrated by Fig. 24.

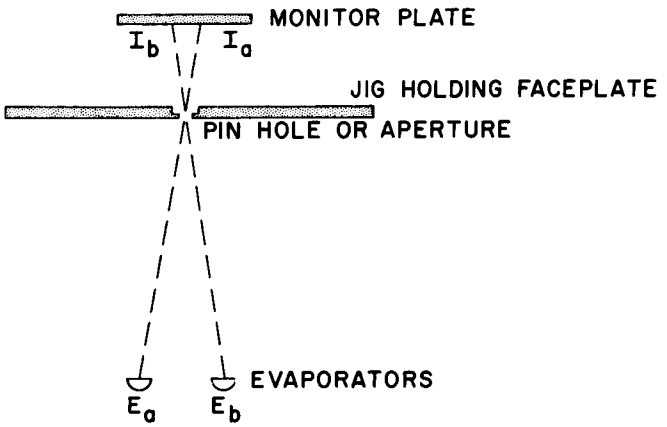


Figure 23. Pin-Hole Evaporation Camera (PHEC)

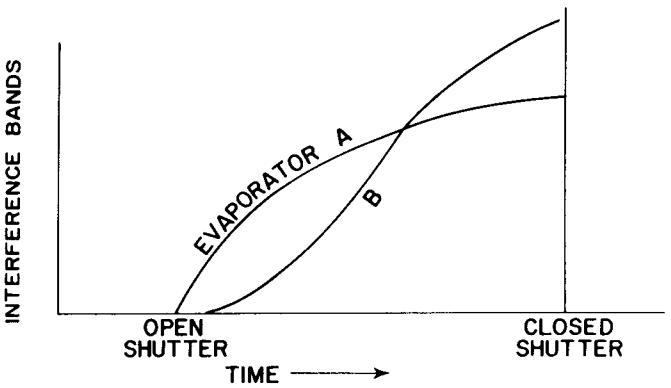


Figure 24. Band Sequence vs. Time for the Evaporation of One Type of Composition-Graded Photolayer

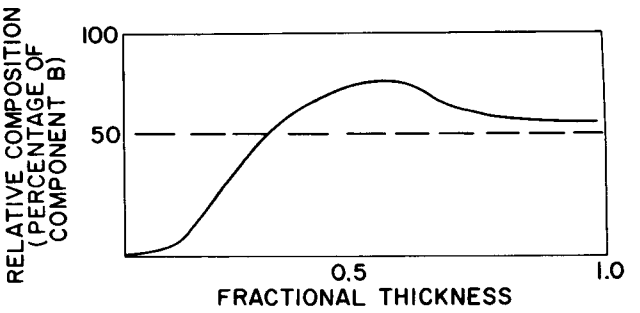


Figure 25. Composition vs. Fractional Thickness for Evaporation Schedule of Fig. 24

PHEC monitoring is not limited to two-component systems, but may be employed for any high-vacuum composition-graded evaporation, regardless of the number of components. "Focal distance," inter-evaporator distance, and aperture size may all be varied to obtain optimum monitoring for any set of component evaporants. The wavelength of the monitoring light is also an important parameter.

REFERENCES

1. Fink, D. G., Television Engineering Handbook, pp. 5-71 to 5-77, McGraw-Hill, New York, 1957
2. Weimer, P. K., S. V. Forgue, and R. R. Goodrich, "The Vidicon," RCA Review, V. 12, No. 3, pp. 306 to 313

Cathodoluminescent Materials

R. N. Summergrad

Lancaster

Luminescence may be defined as the process whereby matter generates nonthermal radiation, that is, luminescence is the production of photons in excess of the photon emission produced by thermal agitation alone. Materials capable of luminescence are called phosphors.

The luminescence process is generally classified according to the means by which the phosphor is excited. For example, the term "photoluminescence" refers to the process whereby phosphors are made to luminesce by the impingement of either ultraviolet or visible radiation upon the phosphor surface. Materials which are excited in this manner find many uses in the lamp industry. The term "electroluminescence" refers to the process by which phosphors emit light when excited by electric fields or currents. Phosphors of this kind are finding increasing usage in such devices as light amplifiers and panel lighting. A third type of luminescence, which is of great importance, is cathodoluminescence which refers to the process of producing luminescence by an electron beam. Cathodoluminescent phosphors are used in such devices as television picture tubes, radar tubes, and oscilloscope tubes. The types of luminescence mentioned above are the most important to the tube engineer. For the sake of completeness, however, one should mention roentgenoluminescence (excitation by high-energy photons, i.e. X-rays and gamma rays), ionoluminescence (excitation by ions), triboluminescence (excitation due to mechanical disruption of crystals), and chemiluminescence (excitation by energy from chemical reactions).

This chapter will be limited to a discussion of cathodoluminescent phosphors. The discussion starts with the presentation of a simple model of the cathodoluminescent process. The next section covers the methods by which phosphors are prepared and illustrates the effect of different conditions of preparation upon the nature of the resultant phosphor. The last section discusses the characteristics of phosphor materials which might affect their utilizability in cathode-ray tubes. Such factors as (1) the stability of phosphors, (2) decay characteristics of phosphors, (3) phosphor body color, (4) color of the emitted light, (5) phosphor efficiency, (6) the variation of light output with current density or accelerating potential, and (7) the variation of light output with temperature are important in deciding whether an individual phosphor may be used in a particular cathode-ray-tube application.

THE CATHODOLUMINESCENT PROCESS

The phosphors that are used in cathode-ray tubes are inorganic solids which may be classified as either semiconductors or insulators. When high-velocity electrons strike the phosphor, the kinetic energy of these electrons is transformed, in part, to light energy. Generally, the phosphors of interest emit in the near-ultraviolet region or visible region of the spectrum.

To understand what occurs between the impingement of an electron beam upon the phosphor surface and the emission of light by the phosphor, the use of a simplified energy level diagram is most useful. Fig. 1* shows the energy levels permitted in a phosphor and some of the electronic transitions which may take place under excitation by an electron beam. The energies of the electrons present in the insulator are plotted along the ordinate, and X, the distance along a crystal axis, is plotted along the abscissa. In a perfect insulator at zero degrees Kelvin the electrons are only allowed to occupy the energy levels designated as filled bands, the highest of which is known as the valence band. The electrons in completely filled bands cannot contribute to the conductivity of the insulator. The highest permitted energy levels for electrons in a solid are found in the conduction band. This band, which is empty at absolute zero temperature, normally has a small fraction of its energy states occupied at room temperature. These electrons may move throughout the crystal and may contribute to the electrical conductivity. Separating the permitted bands are the forbidden regions. In a perfect solid we would find no electrons with energies corresponding to those of the forbidden bands. However, because of the non-ideality of the materials used, electrons are found in localized centers in this region (some localized centers, such as luminescence centers, are deliberately incorporated in a phosphor). Electrons in these centers, such as the trapping center shown in Fig. 1, are localized in space, which means that an electron in such a center is not free to travel through the crystal without first passing into the conduction band.

Consider now the cathodoluminescent process. When an electron strikes a phosphor, it forms a wealth of

*Those who wish to seek further justification for such a plot may refer to the texts listed in Refs. 1 and 2.

secondary electrons along its path by transferring bits of its energy to electrons located in the filled bands. As is shown by the transitions indicated in steps (1) and (2) in Fig. 1, the excited secondaries may eventually find themselves in the conduction band. The removal of these electrons from the filled bands leaves vacancies or "holes" in these bands. For the solid to return to its lowest energy state, the electrons and holes must recombine. An excited electron has several choices available to it, that is, it may return to the ground by any one of a series of competitive processes.

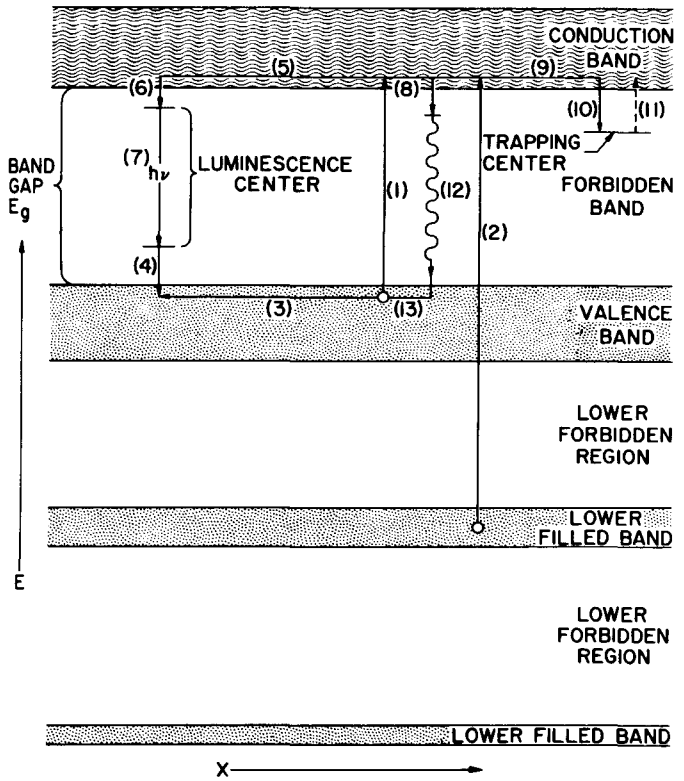


Figure 1. Plot of Electron Energy E in a Phosphor as a Function of Distance Along a Crystal Axis X with electronic Transitions under Cathode-Ray Excitation

1. It may move to a region of the crystal in which an empty luminescence center is located and drop into this center with the subsequent emission of light. This process is shown in Fig. 1, where the "hole" moves through the valence band (3) and is captured by the luminescence center** (4) leaving the center vacant. An electron from the conduction band, which finds itself in the region of the empty luminescence center (5), may then fall into an excited state of the luminescence center (6) with the subsequent release of energy in the form of a quantum of light, $h\nu$ (7) (h is Planck's constant and ν is the frequency of the emitted light). In some phosphors the electron may fall directly from the conduction band to the ground state (or lowest energy state) of the luminescence center, combining (6) and (7) in one step.
2. The electron in the conduction band may undergo non-radiative transitions, as shown in (8), (12), and (13), finally recombining with a hole in the valence band. The nature of the centers in which nonradiative transitions occur is very little understood. In some cases, the luminescence center may itself allow non-radiative transitions to occur within the center, along with the light-emitting transitions[3].
3. The electron may drop into a trapping center (9) (10) from which it will eventually be returned to the conduction band by thermal agitation (11) and will then be in a position to undergo the radiative or non-radiative transtions already discussed.

**This diagram is merely another way of saying that the electron which occupies the luminescence center drops into the vacancy in the valence band.

It is apparent from the foregoing discussion that the emission characteristics of the phosphor will be determined to a large extent by the nature of the centers which are found in the forbidden region between the conduction and the valence band. For example, the efficiency of a phosphor (i.e. the percentage of the beam energy which is converted to light energy) under cathode-ray excitation will largely depend upon the competition between the nonradiative centers and the radiative centers for holes and electrons. The decay characteristics will be controlled by the nature of the luminescence center and the nature of the traps present. The color of the emitted light will depend upon the nature of the luminescence center, while the saturation characteristics will be controlled by the number of luminescent centers per unit volume and the types of center present. Other factors which play an important role in determining phosphor characteristics are the width of the band gap E_g (Fig. 1), secondary emission, and the absorption of emitted light by the phosphor itself.

Excitation by means of particles is an inefficient process because excitation does not take place directly in the luminescence center and, therefore, the excited electron has the possibility of its energy being degraded by means other than by the emission of visible radiation. The picture of the cathodoluminescent process given above is an extremely simplified one. Such steps as the emission of X-rays and short ultraviolet radiation, which in turn might be used to excite the phosphor, have been omitted from the previous discussion. It must be truthfully stated that the detailed picture of the cathodoluminescent process is not available at present and the explanation is extremely difficult to obtain.

PREPARATION OF PHOSPHORS

It is not difficult to prepare cathodoluminescent phosphors, and even the novice may very easily be successful in preparing a new phosphor. The job of the phosphor chemist is a much more difficult one, however, for he must prepare useful new phosphors, that is, phosphors which have a practical application. There is an enormous gap between the rather idealized theoretical picture given in this chapter and the actual preparation of a new phosphor material. At present, aside from those systems which are only slight modifications of known systems, it is not possible to predict the properties of new phosphor systems with any accuracy. Although theoretical calculations which predict the emission characteristics of a simple phosphor system have been made,[4,5] this approach is of very little value in solving the practical problem of preparing new and better materials. A detailed theoretical approach of this type is not useful at the present time because (1) the calculations, even for a simple system, are very time-consuming and the phosphor systems normally used are

much more complicated than the ones for which calculations have been made, and (2) the efficiency of such a system under cathode-ray excitation is impossible to determine theoretically with our present knowledge. The chemist, therefore, must rely upon his chemical knowledge, his intuition, and an approach in which he prepares a large number of samples, using rules of thumb which may help or hinder him in his work.

The most efficient cathodoluminescent phosphors are well-crystallized, white (or slightly colored), high-melting materials which are generally prepared from highly purified "base" materials to which controlled amounts of known impurities, or "activators," are added. There are, however, some phosphors to which no known impurities have been introduced. These, so-called, "self-activated" materials are in many cases thought to be defect structures (i.e., materials which have a slight excess of one type of atom over another). For example, unactivated zinc sulfide is thought to have a slightly greater amount of zinc than of sulfur.

Normally, phosphor preparation is accomplished by high-temperature synthesis; most preparations take place in the temperature range between 700 and 1600 C. High firing temperatures are required to form well-crystallized materials and to incorporate the impurities (i.e., activators) into the crystal lattice.

Phosphor synthesis and phosphor research should be conducted with highly purified materials in an atmosphere free of contamination. Trace amounts of certain impurities may change the characteristics of some phosphors. For example, as little as 2×10^{-5} per cent of nickel in a copper-activated zinc sulfide phosphor will modify the decay characteristics of this phosphor, while, in a zinc sulfide phosphor activated with silver, the incorporation of 10^{-2} per cent of nickel almost completely suppresses the luminescence[6].

The techniques used to prepare phosphors for use in cathode-ray tubes vary with the phosphor system. The properties of the phosphors will depend upon such factors as the base materials (crystal structure, ion size, and bond type of the base material are all factors which affect luminescence), the type of activator or co-activator or both, the firing temperature and time, the atmosphere in which the phosphor is prepared, and the chemical nature of the materials themselves. Obviously, the method of preparation will depend upon the system under consideration and will vary from phosphor system to phosphor system.

The methods of phosphor preparation will be illustrated by examining the procedures used in making several standard phosphors for color TV*. The W-2 phosphor (Zn_2SiO_4:0.012Mn**) is prepared in the following manner:

To two moles of a pure grade of zinc oxide is added 1.2 moles of a "luminescent grade" of silicon dioxide and 0.012 moles of manganese carbonate. These materials are intimately mixed and ground for several hours in distilled water in a thoroughly cleaned pebble mill. After the water has been removed by filtration, the material is dried and then fired in large, open silica boats at 1250 C for two hours, in an atmosphere of air. Next, the phosphor is cooled and further processed to fit the demands of the particular method of application which will be used to form the phosphor screen.

Because the ZnS:0.015% Ag:Cl phosphor also used in the color TV tube is much more sensitive to very small amounts of contaminants (such as nickel, iron, and copper) than the Zn_2SiO_4:Mn phosphor, greater precautions must be taken in the preparation of the sulfide phosphor.

It seems to be generally true that the smaller the amount of activator required to produce efficient luminescence, the more easily is the luminescence affected by impurities. The zinc sulfide base to be used is precipitated from a purified zinc sulfate solution. The solution may be purified by partial precipitation of zinc sulfide. The partial precipitate contains such impurities as silver, nickel, iron, cobalt, and copper, the sulfides of which are much less soluble than zinc sulfide. The solid sulfide containing the impurities is removed and the precipitation of pure zinc sulfide is completed. About 0.015 per cent by weight of silver is then added as silver nitrate together with a "flux" in the form of purified sodium chloride. An intimate mixture may be prepared by milling the ingredients in a pebble mill. The resultant phosphor mix is dried and then fired in tightly covered containers in air at 800 C for two hours. Next, the surface of the phosphor charge must be cleaned to remove the portion that has been oxidized to zinc oxide. Further processing may be required in order to prepare the phosphor for its final application.

The phosphor syntheses discussed above illustrate several important points about phosphor preparation:

1. Phosphors of the oxide type generally require larger amounts of activator than do the sulfide phosphors. Along with this need, the purity requirements of oxide systems are not as stringent as sulfide systems.
2. All activated phosphors require intimate mixing of the solid starting materials to insure uniformity of the final product. For the zinc-silicate phosphor activated with manganese, the intimacy of mixing is required to insure the production of a uniform base material, as well as to insure the uniform distribution of the activator in the base matrix, while the blue-emitting sulfide preparation requires only that the activator be uniformly distributed.

*See Table I and figures for the characteristics of these phosphors.

**This symbolism will be used throughout the text. The first chemical formula, that of zinc silicate, refers to the "base" material (the material into whose structure we are placing the activator). The activator symbol (Mn for manganese in this case) and the number of moles of activator per mole of base material follows the chemical symbol for the base material. Where percentages are used, they will refer to weight per cent. In some cases, co-activators will be listed following the activator.

3. Many oxide phosphors are found to be most efficient when the composition of the base material deviates from stoichiometry to some extent. This statement is found to be true of the Zn_2SiO_4:Mn phosphor, where a slight excess of silica is used to optimize the efficiency. Fig. 2 shows the variation of the efficiency of manganese-activated zinc silicate (willemite) as a function of the composition of the base material.

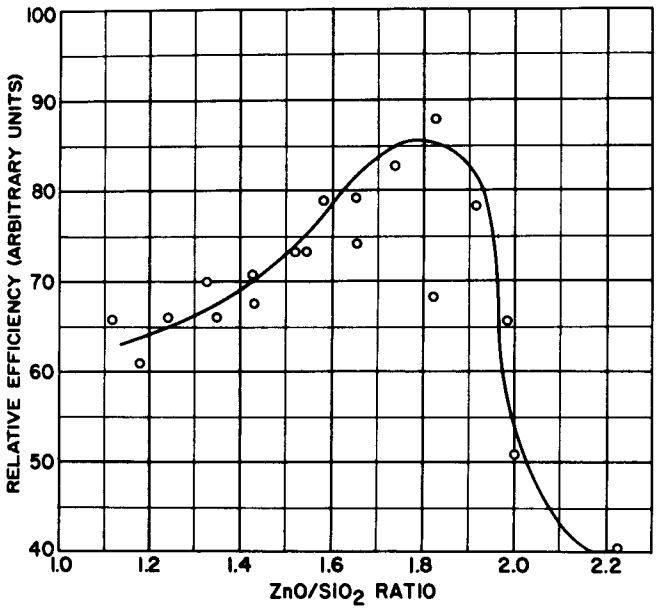


Figure 2. The Effect of the $ZnO:SiO_2$ Ratio on the Luminescent Efficiency of Willemite (after M. R. Royce)

4. In many sulfide phosphors activated with monovalent positive ions such as Ag^+ or Cu^+, a "flux" such as NaCl is added to the phosphor mixture. This flux serves two purposes. It is usually a low melting material which is thought to partially dissolve and reprecipitate the phosphor base material and, therefore, to promote crystallization. Its second function appears to be to facilitate the entry of the monovalent positive ion (Ag^+ in this case) into the ZnS lattice where it substitutes for a doubly charged Zn^{++} ion. It does this by substituting a singly charged, negative Cl^- ion for a doubly charged, negative $S^=$ ion. This substitution allows the crystal to remain electrically neutral. The Cl^- ion is called a "co-activator"; the phenomenon described is known as charge compensation. Another way that charge compensation may occur is by the formation of lattice vacancies. For example, the substitution of two singly charged positive ions for two doubly charged positive ions may be compensated for by a negative-ion vacancy. The second type of charge compensation is generally less favorable energetically than the first.

Fig. 3 illustrates some of the different types of defects and substitutions which may be found in a phosphor crystal of the type under discussion. The crystal lattice shown in this diagram is an idealized planar lattice; real phosphor crystals have three-dimensional lattices.

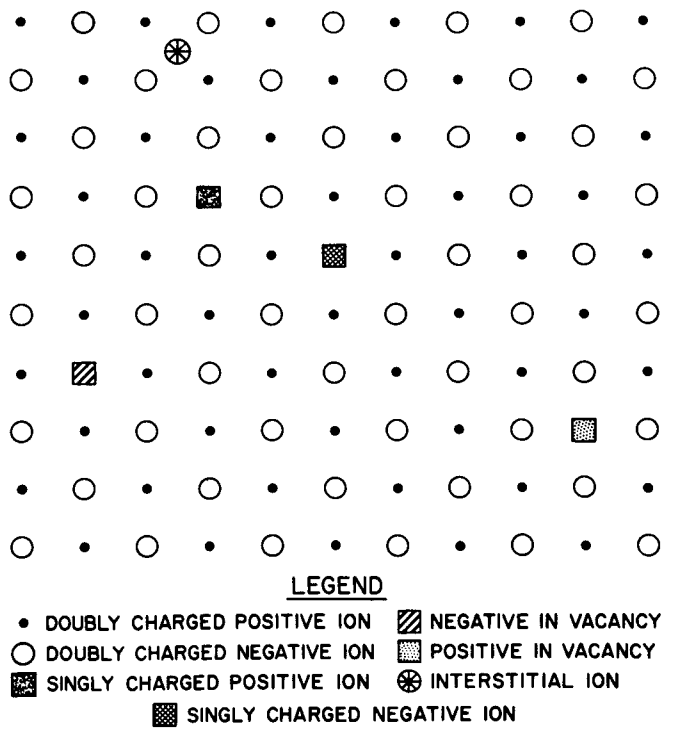


Figure 3. Idealized Planar Lattice Containing Doubly Charged Negative and Positive Ions with Singly Charged Ions as Impurities

PHOSPHOR RESEARCH

The search for new or improved phosphors may be classified under the following headings:

1. The formation and activation of new base materials (i. e., base materials heretofore unknown).
2. The activation of established compounds with new activators or with previously tried activators in new ways.
3. The modification of established phosphors by means of preparatory techniques or impurity additions.

It is sometimes difficult to sharply distinguish between these classifications.

Although a complete review of the preparation of new phosphors is beyond the scope of this chapter, a brief discussion of some of the methods used in preparing new materials and in modifying the characteristics of old ones will be given.

The phosphor chemist, in search of a new base material, chooses a group of starting materials that he feels has the possibility of forming new compounds. His choice may be completely a matter of guesswork or he may take into consideration such factors as the chemical nature of the materials involved, the sizes of the cations and anions involved, and the probability that a white material will be formed. In spite of all the consideration given to the factors involved, it is still extremely difficult, if not impossible, to predict whether

compound formation will take place in a new system. After he has chosen his materials, he attempts to form compounds by reacting starting mixtures, which are varied over the entire composition range, under what he considers to be the proper conditions*. Compound formation is then determined by X-ray diffraction. Initially, large increments in composition are taken to indicate whether new compounds are being formed. If new compound formation is indicated (by the appearance of lines in the diffraction pattern other than those produced by the starting materials), then he is justified in taking smaller increments about the composition which shows the most intense lines of the new compound. In this way, the exact composition may be established. He may then attempt to activate this material.

Normally, when a new phosphor system is found, a great deal of work must be done to optimize the system in order to produce a phosphor with the most desirable characteristics attainable. Such things as phosphor efficiency, color, decay characteristics, saturation characteristics, and temperature dependence might all need modification. To affect these phosphor properties, the chemist may modify the preparatory conditions (independently or in conjunction with one another) as follows:

1. By varying the reaction temperature or reaction time or both.
2. By changing the atmosphere in which the preparation takes place.
3. By varying the activator concentration.
4. By changing the composition of the base material.
5. By fluxing the mix.
6. By adding impurities.

Variations in firing temperature, or firing time, or both may affect the completeness of reaction, the distribution of activator, the crystallinity of the phosphor, and the crystal structure of the base material. All of these factors may change the properties of the phosphor. For example, heating the ZnS:Ag phosphor to temperatures greater than 1020 C causes the crystal structure to change from the low-temperature cubic form to the high-temperature hexagonal form. The crystal change is responsible for a shift in the emission to shorter wavelengths.

In most cases, changing the atmosphere in which a phosphor is prepared will affect the characteristics of the phosphor. The extent of the effect will depend upon the phosphor system under consideration. In many aluminates, such as $MgAl_2O_4$:Mn, firing in a reducing atmosphere (e.g., H_2 gas) produces a green-emitting phosphor, while firing in an oxidizing atmosphere (e.g., air) produces a red-emitting phosphor. This change in emission color is due to the oxidation of manganese to a higher valence state under oxidizing conditions.

*The proper conditions for compound formation may be very difficult to ascertain. For example, the two-component system, CdO and B_2O_3, was only recently worked out completely [7,8], although the system has been known and studied for many years. The determination of the proper preparatory conditions may be greatly aided by the use of "differential thermal analysis" and the use of a high-temperature balance.

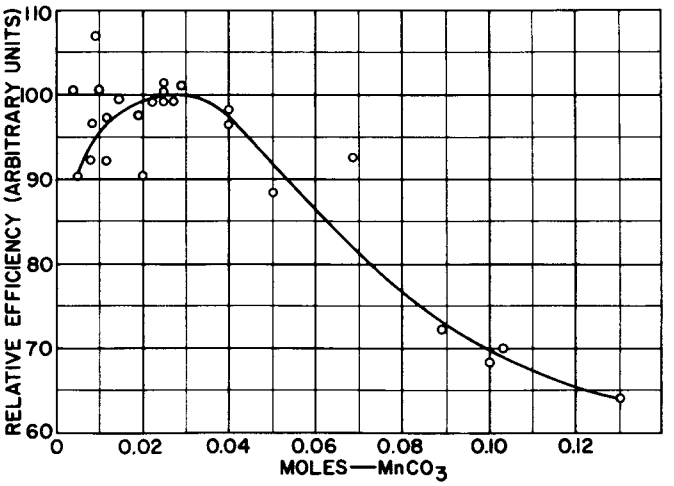


Figure 4. The Dependence of Luminescent Efficiency upon Activator Concentration in Zinc Silicate: Manganese

The variations of the activator concentration of a phosphor will affect its efficiency, may cause a marked change in the "temperature breakpoint," cause a slight shift in emission color, and may affect its decay characteristics. The effect of manganese activator concentration upon the efficiency and the decay of the Zn_2SiO_4: Mn green phosphor is shown in Figs. 4 and 5, taken from the work of M. R. Royce.

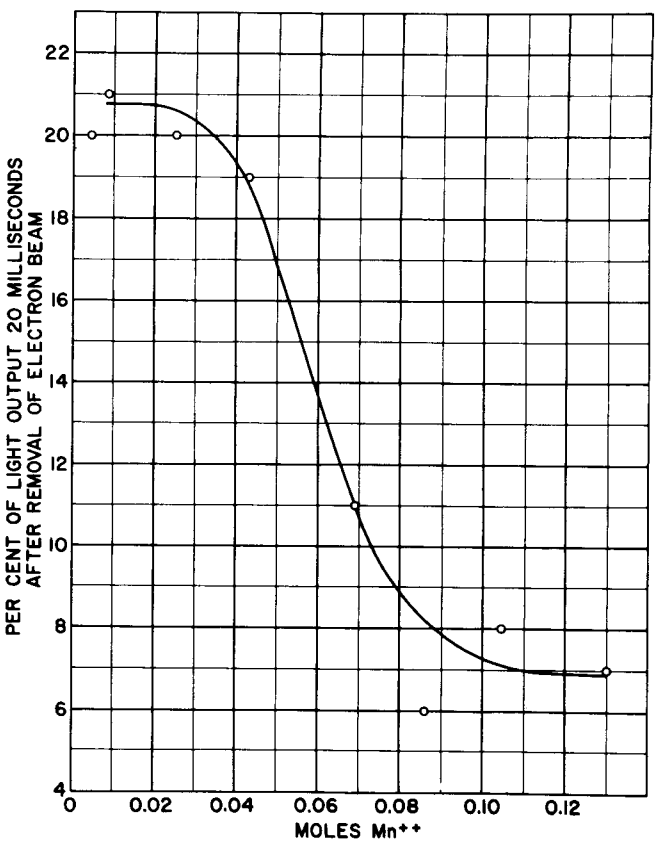


Figure 5. The Effect of Manganese Concentration Upon Persistence in Zn_2SiO_4:Mn

The addition of impurities to an established phosphor, the addition of a flux, and the variation of the composition of the base material upon the phosphor characteristics were discussed and illustrated in the previous section.

PHOSPHOR CHARACTERISTICS

Proper design of cathode-ray tubes must take into account the inherent properties (and very often limitations) of the phosphors used for the screen. Among the many factors that must be considered are the following:

1. Optical Properties: Amount and type of light necessary, spectral distribution, visual efficiency decay characteristics, current-saturation properties, effect of accelerating potential on the phosphor emission, and temperature stability under operating conditions.
2. Physical Properties: Particle size of the phosphor, stability of the phosphor to tube processing, and resistance of the phosphor to "burning" during tube operation.

It is rare when all the desirable properties demanded can be obtained. Most often, the design engineer must accept phosphors which do not completely satisfy all his requirements because the phosphor chemist can produce nothing better. Table I lists several RCA phosphors, indicates their chemical composition, and tabulates some properties which might be of interest to the design engineer. Figs. 6, 7, and 8 show the spectral energy curves of these phosphors. For a more complete list, see form No. TPM-1508, entitled "RCA Phosphors." An engineer's choice of phosphors is not limited to the partial list shown here, or even to the more complete list on "RCA Phosphors," since many "experimental" phosphors, which at present have no commercial application, are available.

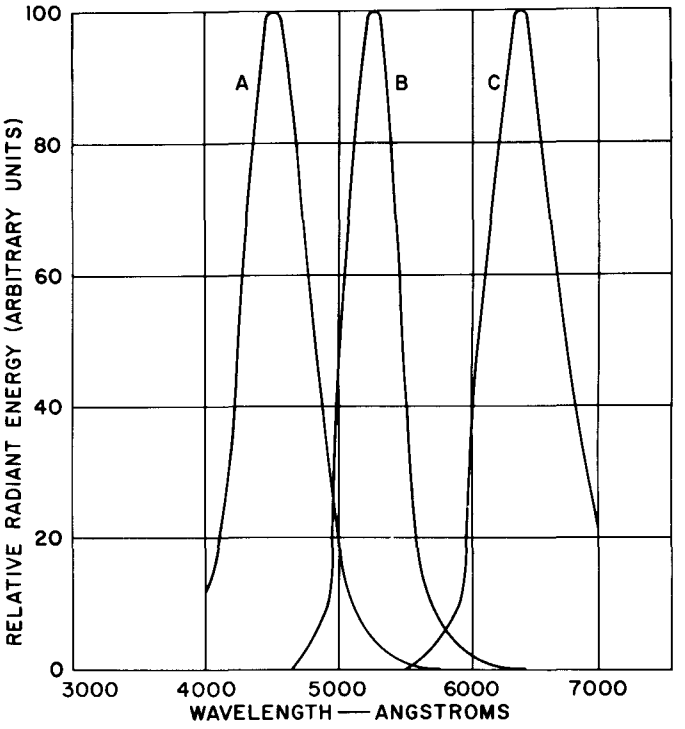


Figure 6. Relative Radiant Energy as a Function of Wavelength for RCA Phosphors (A) 33-Z-265; (B) 33-W-2A, (C) 33-Z-639D. Excitation voltage =8 kilovolts; current density =1.5 μA/CM² for defocused spot; ordinates for curves A, B, and C are independent

Phosphor Efficiency

From the previous discussion of the mechanism of cathodoluminescence it might be expected that the efficiency of this process would be rather poor. One is, therefore, somewhat surprised by the efficiencies of

Table I

Characteristics of RCA Phosphors

Phosphor Formula	Emission Color	Persistence Per cent of initial brightness at time interval	Wavelength of Peak Radiant Energy angstroms	C.I.E. Coordinates x	C.I.E. Coordinates y	Relative Radiant Energy* Total	Relative Radiant Energy* Visual	RCA Number
Zn_2SiO_4:Mn	Green	Medium: 18 at 20 milliseconds	5250	0.203	0.728	100	100	33-W2-A
ZnS:Ag	Blue	Short: 0.1 at 20 milliseconds	4500	0.146	0.52	242	26	33-Z-265
ZnO:Zn	Ultraviolet	Extremely short: 10 at 0.05 millisecond	3910	----	----	---	---	33Z-613
	Green	10 at 3 milliseconds	5100	0.246	0.439	85	59	
ZnCdS:Cu	Greenish-yellow	Long	5350	0.291	0.546	118	104	33Z-601
$Zn_3(PO_4)_2$:Mn	Red	Medium: 29 at 20 milliseconds	6380	0.674	0.326	73	22.5	33Z-639D

*Relative to 33-W2-A

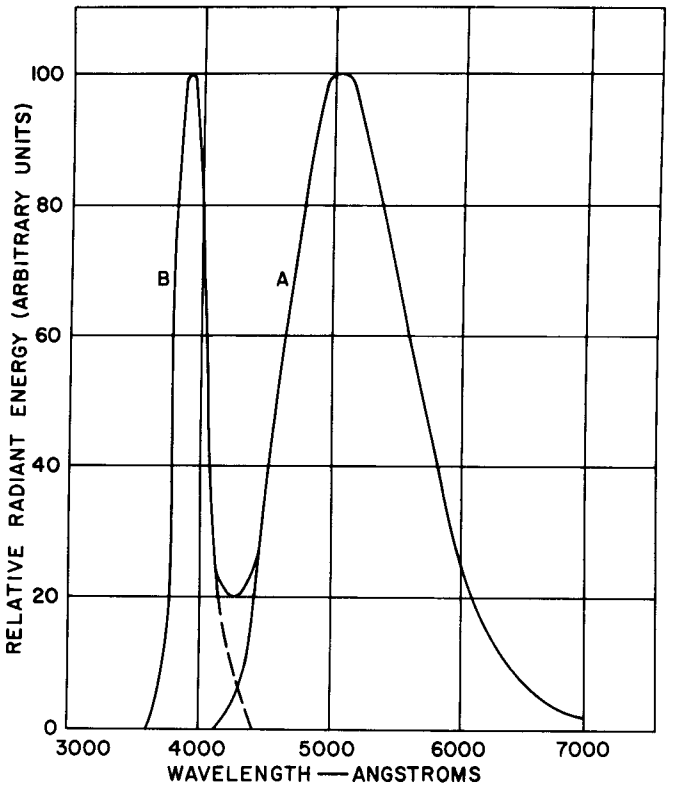


Figure 7. Spectral Emission Curves for Zn): Zn (33--Z--613) Showing the Emission Change When Current Density Is Varied at an Excitation Voltage of 8 Kilovolts: curve A, 1.5 μA/cm²; curve B, 100 uA/cm²

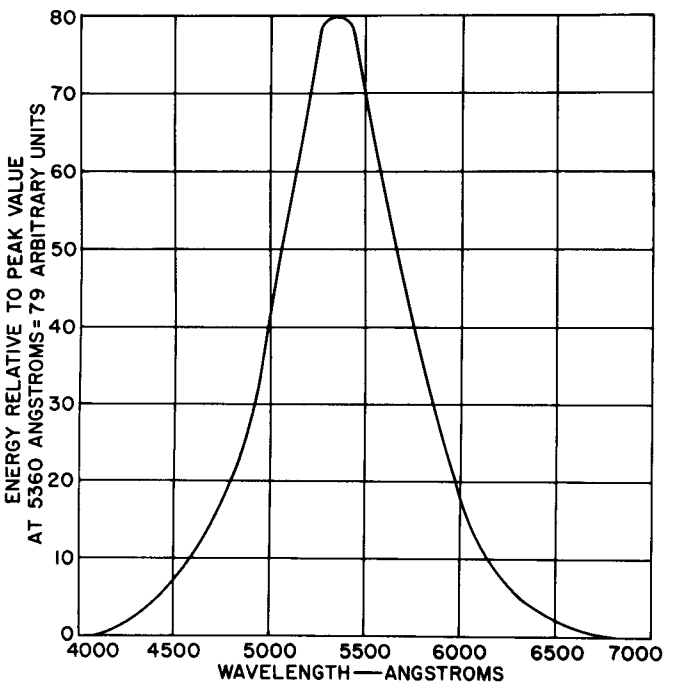


Figure 8. The Spectral Emission Curve of ZnCdS:Cu (Z-601)

some of the better cathodoluminescence materials as measured by Bril and Klasens[9, 10] (see Table II). Note the great disparity between the efficiencies of the phosphors in the sulfide systems activated with "monovalent" activators, such as silver and copper, and the efficiencies of the phosphors with oxides as the base material, or the efficiencies of the manganese-activated sulfides. There is no known reason for this difference in efficiency, but all known measurements of phosphor efficiency, although they may differ from the values listed in Table II, show the same relative differences[6] between these two different phosphor types.

Table II

Intrinsic Efficiencies

Phosphor Formula	% Energy Conversion*
$ZnS:10^{-4}Ag:10^{-4}Al$	25
(50% ZnS - 50% CdS):10^{-4}Ag:10^{-4}Al	19.5
$ZnS:10^{-5}Cu:10^{-4}Al$	25
ZnS:0.015Mn	4.5
Zn_2Si0_4-0.01Ti	8.5
Zn_2Si0_4- 0.004 Mn	8.5
$Cd_2B_2O_5$:0.002Mn	5
$Zn_3(PO_4)_2$:0.03Mn	5.5

*% Energy conversion refers to the percentage of the electron-beam energy which is converted to light energy.

Decay Characteristics of Phosphors

In most cathode-ray tubes the excitation of the screen is intermittent. Because of this, the characteristics of the rise and decay of luminescence are of interest. For example, in television tubes the initial rise and decay of the emission during the first few milliseconds are of interest, while for radar applications the growth and decay curves for periods of up to several seconds must be studied. Several papers[11, 12] show that the growth and decay of cathodoluminescence may be divided into an initial process and later processes. In many phosphors the initial process increases or decays exponentially with time, while the later stages of the rise or the decay of luminescence, for most phosphors, do not follow an exponential law. In almost all cases, the simple exponential portion of the decay curve remains almost completely unchanged when the phosphor temperature or the current density of the electron beam are varied. In contrast to this behavior, a phosphor which suffers a power-law decay (i.e., $L(t) \propto e^{-n}$, where n is a constant, t is the time after excitation has ceased, and L(t) is the light output at time t) has its decay characteristics markedly changed by either the temperature or the current density. The effects of temperature and current density upon the decay of a Zn_2Si0_4:Mn phosphor are shown in Figs. 9 and 10. The exponential portion of these decay curves ($L(t) \propto$

e^{-at} where a is a constant) indicates that a monomolecular process is the initial rate-controlling step. The first-order kinetics is thought to be due to the electronic transition from the excited state in the manganese center to the ground state of the center. The power-law portion of the curve is thought to result from the presence of traps whose depth and concentration control the rate of decay (see (11) Fig. 1). The foregoing theory has been confirmed by experimental observation. Since the nature of the luminescence center is very little affected by current density or temperature, one would not expect the exponential portion of the decay curve to be changed much by change in either current density or temperature. However, the occupancy of traps and the rate at which they are emptied depends upon the temperature and the current density, therefore, the power-law portion of the decay curve should be affected markedly by both the temperature and the current density. Figs. 9 and 10 offer evidence which favors the suggested explanation. Phosphors of the ZnCdS:Cu type owe their long persistence to the presence of deep traps. The whole decay curve of such phosphors may follow a power law, as shown in Fig. 11.

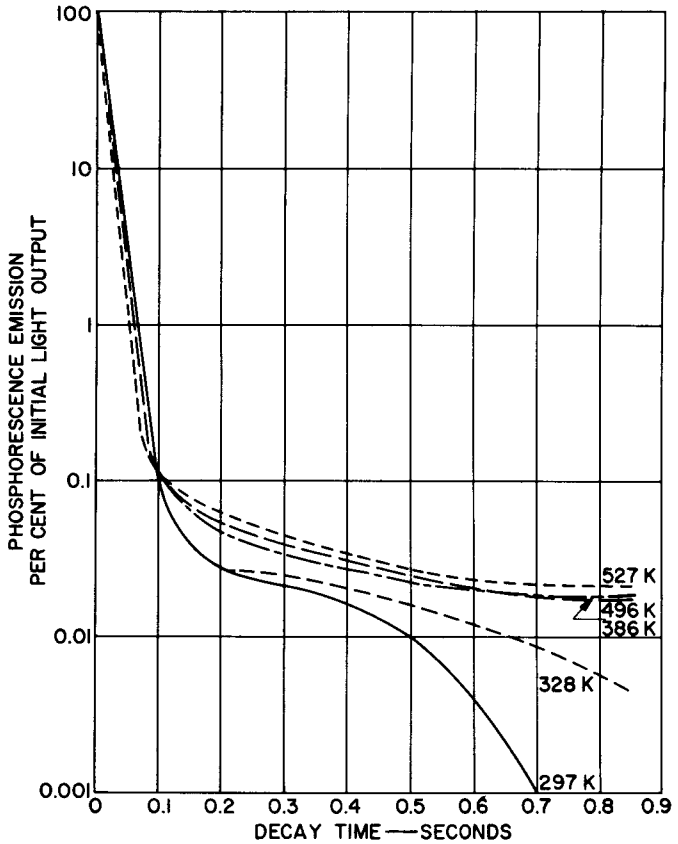


Figure 9. Phosphorescence Emission as a Function of Decay Time for Various Temperatures (after Leverenz)

The shortest decays (among the efficient phosphors) are found in ZnO:(Zn), cubic MgS:Sb, etc., which decay to half their peak intensities in times of the order of 10^{-6} seconds. The longest, efficient, exponential decay is found in ZnF_2:Mn-type systems where the exponential component is still discernible about one second after the exciting source is removed. The photoluminescence of such t^{-n} type phosphors as ZnS:Cu is still discernible after several days.

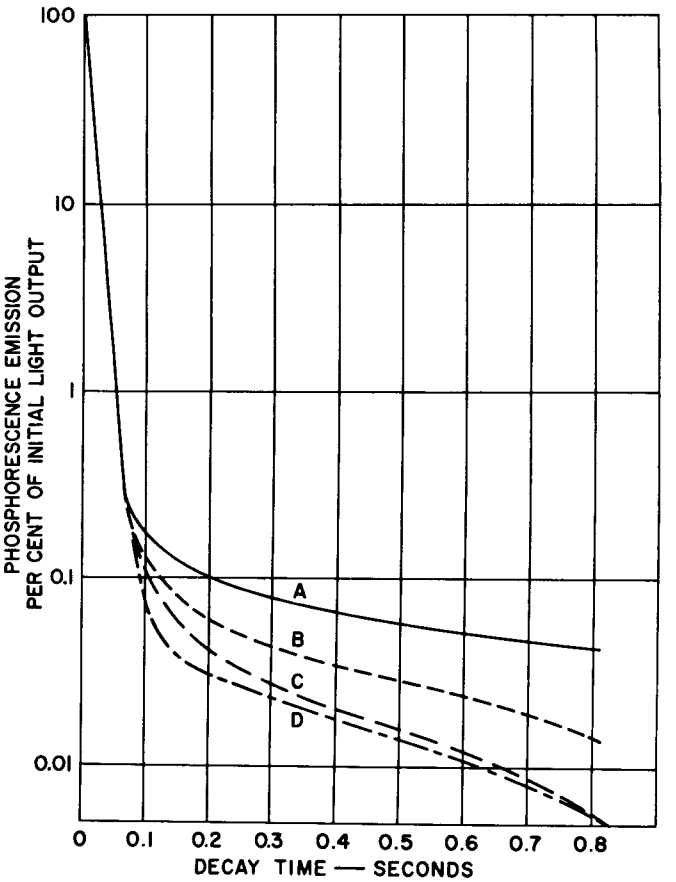


Figure 10. Phosphorescence Emission as a Function of Decay Time for Various Excitation Current Densities. Excitation current density: curve A, 0.04 μA/cm²; curve B, 0.25 μA/cm²; curve C, 0.56 μA/cm²; curve D, 1.5 μA/cm²; (after Leverenz)

Those readers who are interested in a more detailed discussion of phosphor decay are referred to Ref. 6.

The Variation of Light Output with Accelerating Potential and with Current Density

It is often desirable for a phosphor to show a linear rise in light output with current density and with accelerating potential. In some cases, if linearity is desired, the nonlinearity may be compensated electronically.

Accelerating Potential. The light output of a phosphor screen as a function of accelerating voltage depends upon the nature of the phosphor screen and the way that the phosphor has been deposited. The light output of a screen as a function of voltage is given by the following equation: $L = K(i)(V-V_0)^q$, where L is the light output, V is the accelerating potential, V_0 is the "dead" voltage, K(i) is a constant which is dependent on the current density, and q is a constant which ranges from 1 to 3 depending upon the screen preparation. The dead voltage V_0 is thought to be due to surface contamination. Fig. 12, taken from the work of Strange and Hender-

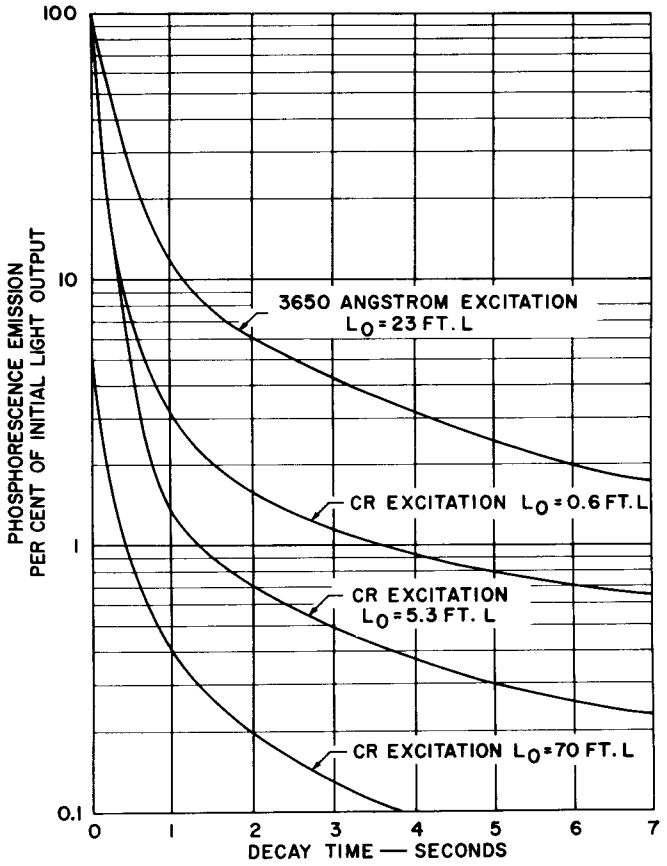


Figure 11. Phosphorescence Emission as a Function of Decay Time for Various Cathode-Ray Excitation Levels (after Leverenz)

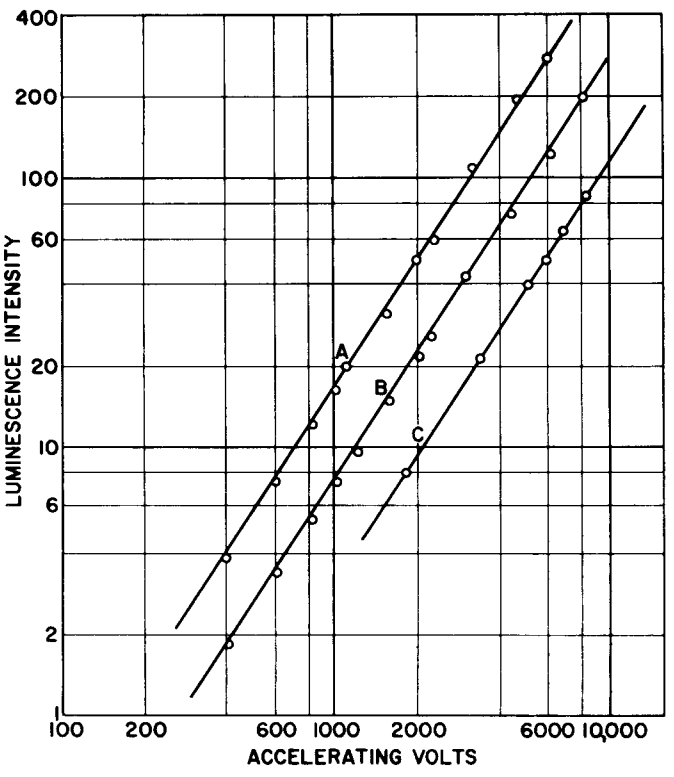


Figure 12. Variation of Cathodoluminescence with Electron Accelerating Potential for a Zn_2SiO_4--Mn Phosphor Excitation current density: curve A, 5 $\mu A/cm^2$; curve B, 100 $\mu A/cm^2$; curve C, 550 $\mu A/cm^2$

son[11], shows the effect of accelerating potential on the light output for the green, manganese-activated zinc silicate phosphor screen.

Current Density. The luminescence intensity per unit area is directly proportional to the current density at low values of the current density, but saturation occurs at higher values. At constant excitation, the current density at which saturation begins is a characteristic of the phosphor material. As a rough rule of thumb, under intermittent excitation the phosphors with the greatest number of luminescence centers per unit volume show the least saturation. Fig. 13 illustrates this point. The Zn_2Si0_4:Mn phosphor has several orders of magnitude more activator than the ZnS:Ag phosphors which show much less saturation.

Temperature: Its Effect on Luminescence

The efficiency of phosphors is fairly constant up to a certain temperature and decreases rather rapidly thereafter. The temperature at which this decrease in efficiency occurs is called the temperature breakpoint. The temperature breakpoint is a characteristic of the phosphor, but may also depend upon the means of excitation.

Cathode-ray tubes are normally operated at room temperature, so that one would, perhaps, expect that the variation of light output with temperature would be of no practical importance. However, in cases where a great deal of power is concentrated in a very small region of the screen of a cathode-ray tube, the instantaneous phosphor temperature may be appreciably increased. If the temperature breakpoint of a phosphor is only slightly above room temperature, its efficiency may drop appreciably under high voltage and high current density conditions. Such a drop in efficiency would be caused by both the saturation of the phosphor and the increase of the temperature beyond the temperature breakpoint.

The effect of the temperature breakpoint upon phosphor efficiency may best be illustrated by comparing the blue phosphor, ZnS:0.015%Ag:0.015%Al, which is used in the color projection tube, with the blue phosphors ZnS:0.015%Ag:Cl, used in the color tube. The projection-tube phosphor, which has the higher temperature breakpoint of the two, is less efficient under normal operating conditions, but in the projection tube the situation is sharply reversed.

REFERENCES

1. Shockley, N., Electrons and Holes in Semiconductors, Chapter 5, D. Van Nostrand & Co., Inc., 1950
2. Kroger, F. A., Some Aspects of the Luminescence of Solids, Chapter 1, Elsevier Publishing Company, Inc., 1948

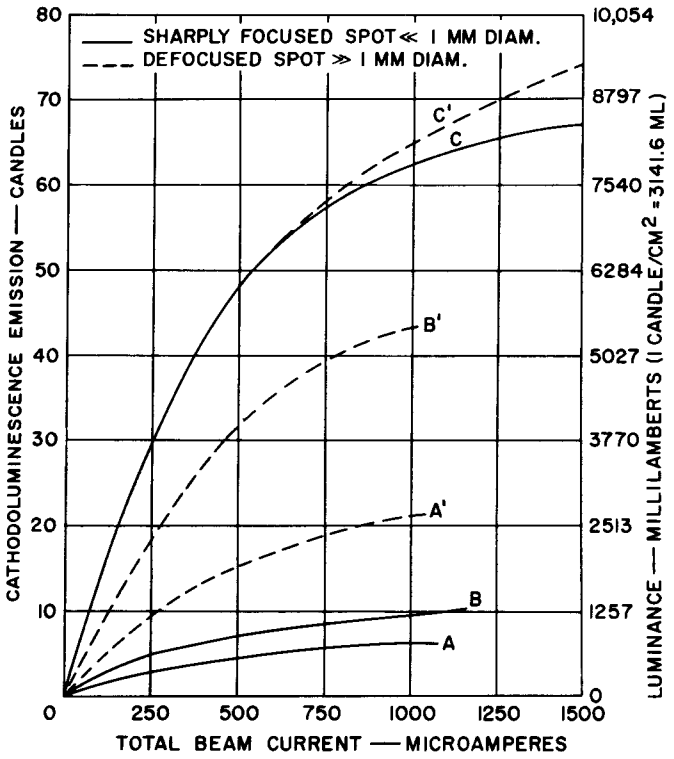


Figure 13. Cathodoluminescence as a Function of Beam Current (at 21 Kilovolts with 5 x 5 cm scanned pattern). Phosphor: curve A, cubic ZnS:Ag (0.0025), 850 C, 7 mg/cm^2; curve B, hexagonal ZnS:Ag (0.015), 1250 C, 8 mg/cm^2; curve C, rhombohedral Zn_2SiO_4:Mn (0.3), 1250 C, 5 mg/cm^2 (after Leverenz)

3. Shulman, J. H., British Journal of Applied Physics, Supplement No. 4, S64, 1955
4. Williams, F. E., J. Chem. Phys., 19, 457, 1951
5. Williams, F. E., J. Chem. Phys., 57, 780, 1953
6. Leverenz, H.W., An Introduction to Luminescence of Solids, John Wiley, and Sons, New York-London, 1950
7. Subbarao, E. C., and F. A. Hummel, J. Electrochem. Soc., 103, 29, 1956
8. Hummel, F. A., and E. C. Subbarao, J. Electrochem. Soc., 104, 619, 1957
9. Bril, A., and H. A. Klasens, Philips Res. Rep., 7, 401, 1952
10. Bril, A., and H. A. Klasens, Philips Res. Rep., 10, 305, 1955
11. Strange, J. W., and S. T. Henderson, Proc. Phys. Soc., 58, 368, 1946
12. Marten, S. T., and L. B. Headrick, J. App. Phys., 10, 116, 1939

Application and Colorimetry of Phosphor Screens

A. E. Hardy

Lancaster

SCREEN APPLICATION TECHNIQUES

Phosphors are used in cathode-ray tubes to convert electron energy into light. The phosphors for this purpose are inorganic crystalline materials with a particle size of 1 to 10 microns. Much effort has been spent in developing methods for applying the phosphor particles in the form of a thin uniform screen to the inner surface of the faceplate of the cathode-ray tube. The amount of phosphor deposited is given in milligrams per square centimeter and ranges from 1.5 to 20. Screen weight per square centimeter for any particular tube type is determined mainly by the particle size of the phosphor which is used and the voltage at which the ultor is operated. Usually, screen weight per square centimeter is increased with increase in ultor voltage. Some special double-layer cascade-type screens for radar applications require heavy screens for good physical separation between the layers. The minimum screen weight per square centimeter is determined largely by the quantity of material required to give complete glass coverage which in turn is related to the particle size of the phosphor.

Screen weights per square centimeter for some typical tubes are given in Table I.

Table I

Tube Type	Ultor Volts	Screen Weight Milligrams per sq. cm.
1EP1	1500	2.0
5FP7A	8000	22.0
10BP4A	12000	5.2
17LP4	16000	5.2
21CEP4*	18000	3.0
21CYP22	25000	2.0-3.0
5TP4*	27000	3.0
7NP4*	80000	8.0

* Aluminum-backed screens

The light output of screens operating at ultor voltages above 6000 volts may be considerably enhanced by a deposit of bright aluminum film on the beam side of the phosphor screen. This film serves as a mirror to increase the light output by a factor as high as 1.8.

Numerous methods have been proposed or used for laying down phosphor screens. Some of these are: water settling, slurrying, dusting, air spraying, liquid spraying, electro-spraying, tacky film, smoking, cataphoresis, monolayer flotation, evaporation, painting, printing, decalcomania transfer, embedding in face plate, and swirling.

APPLYING PHOSPHOR SCREENS FOR BLACK-AND-WHITE PICTURE TUBES AND OSCILLOSCOPE TUBES

Water settling is by far the most widely used technique for applying screens to all tubes except color tubes. This method dates back to at least 1930 and has been used exclusively for black-and-white screens made commercially in the U.S.A.

Although the water-settling method is quite old, many changes and refinements have been made to bring it to its present high level of development. Each tube manufacturer will have one or more standard settling procedures which differ slightly in the concentrations and in the sequence of addition of materials, but all settling schedules are basically the same.

The flow chart shown in Fig. 1 outlines the major steps in screening, filming, and aluminizing a black-and-white television tube.

The phosphor powder, which is actually a mixture of insoluble blue and yellow phosphor crystals, is prepared as a water suspension either with or without ball milling. Ball milling may be required in some instances to break up phosphor crystal aggregates and aid dispersion.

A quantity of cold water, containing an electrolyte such as barium acetate, is added to the tube bulb to produce a cushion layer. A quantity of potassium silicate, which is the screen binder, is added to the cushion water. The phosphor suspension is added after, or with, the potassium silicate by means of a long-stem funnel equipped with a nozzle or large-hole spray tip at the delivery end. The phosphor is then allowed to settle out of the liquid under the influence of gravity. During the settling period, the electrolyte causes a partial gelling of the silicate. The silica gel becomes an adhesive between particles within the screen coating and between the coating and the faceplate. (Particle-to-particle binding is called cohesion while particle-to-glass binding is termed adhesion.)

At the end of the settling period, the cushion layer is

carefully poured off and the screen is forced-air dried. With large-scale production, the settling process is carried out on a conveyor belt. Pour-off occurs as the bulb passes down over the end of the belt. Small-scale production and laboratory settling are frequently done on a table that can be tilted. Siphoning is sometimes used with either belt or table settling. Although water settling is still largely an art, considerable technology has been developed and applied to it.

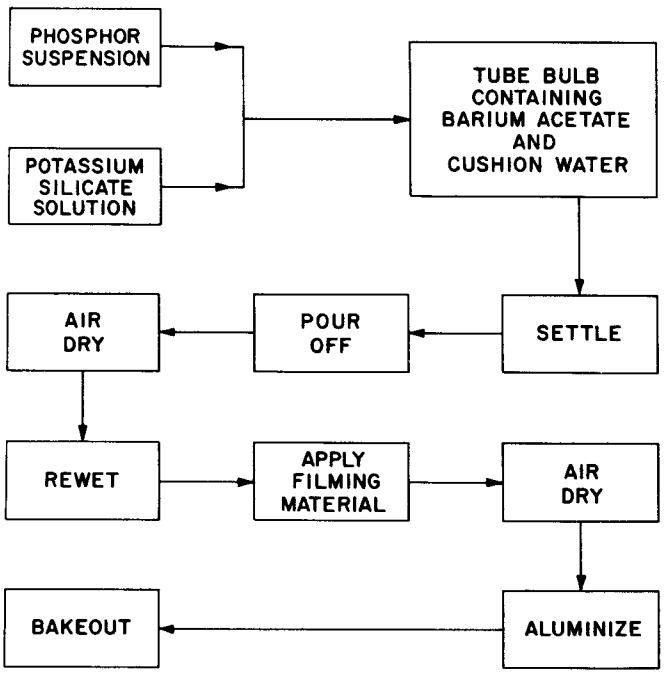


Figure 1. Major Steps in Screening, Filming, and Aluminizing a Black-and-White Television Tube.

STOKE'S LAW

Particles settling through a liquid under the influence of gravity alone quickly reach a terminal velocity which is given by Stoke's Law.

$$V = \frac{2ga^2(d - d_o)}{9\eta}$$

where V = terminal velocity in cm/sec
g = acceleration due to gravity in cm/sec^2
a = radius of particle (a spherical shape is assumed) in cm
d = density of particle in gm/cc
d_o = density of the settling medium in gm/cc
η = coefficient of viscosity in poises

Table II shows the terminal velocity and time required for particles in the 1 to 20 micron range to settle through 10 centimeters of water.

Table II

Particle Diameter-Microns	Terminal Velocity-cm/minute	Time to Settle 10 cm-minutes
1	0.01	1000
2	0.04	250
5	0.25	40
10	1.0	10
20	4.0	2.5

The principle of differential settling rates may be used to make a particle-size separation. The process of fractionating particles by suspending them in a liquid, allowing partial settling to take place, and then pouring off particles still in suspension, is called elutriation.

SCREEN ADHERENCE

Sodium silicate was the first material to be used as a binder in water settling. With this material, standing times of about eight hours were necessary to develop sufficient adherence for pour-off. Later, when sodium sulfate was added as an electrolyte, standing times were reduced to one hour. Changing from the sodium salt to the potassium salt eliminated a form of screen burning which had been prevalent for some time. Still later, it was found that the use of divalent salts, such as barium acetate, further reduced standing times to 10 to 15 minutes. Other electrolytes, such as acetic acid, barium nitrate, strontium nitrate, and potassium bicarbonate, have been used commercially, but the tube industry now uses barium acetate.

The concentrations of silicate and acetate for best adherence have been determined empirically. There is no apparent stoichiometry. Adherence of a settled screen is determined by measuring the size of a hole eroded in the coating by a calibrated jet of water of specified duration. A typical contour diagram which depicts adherence as a function of silicate and acetate concentration is shown in Fig. 2. The unit of adherence is the reciprocal of the eroded-hole diameter in centimeters.

The screen adherence discussed in the previous sections is the adherence that exists during pour-off and is called screen wet strength or wet adherence. The adherence of the dry screen is also important and, therefore, some compromise must be made between wet and dry adherence. In general, dry adherence increases directly with the amount of silicate used. Dry adherence is measured by determining the amount of air pressure required to blow off a portion of the dried screen coating.

In the preparation of a screen for aluminizing, a dry screen is rewet with water. Rewetting the coating leads to still another adherence consideration: that of rewet adherence. It is obvious that the concentrations of silicate and acetate actually used represent a compromise among the requirements for wet adherence, rewet adherence, and dry-screen adherence.

FILMING AND ALUMINIZING BLACK-AND-WHITE SCREENS

Placing a bright aluminum film on the back of a phosphor screen increases light output by a factor of up to 1.8. In addition, the aluminum film improves large-area contrast by minimizing the harmful effects of optical and electron reflections from the inside wall of the tube envelope. A further advantage is the elimination of screen charging at operating voltages above 12,000 volts. A non-aluminized phosphor screen depends on secondary emission to keep the screen at ultor potential.

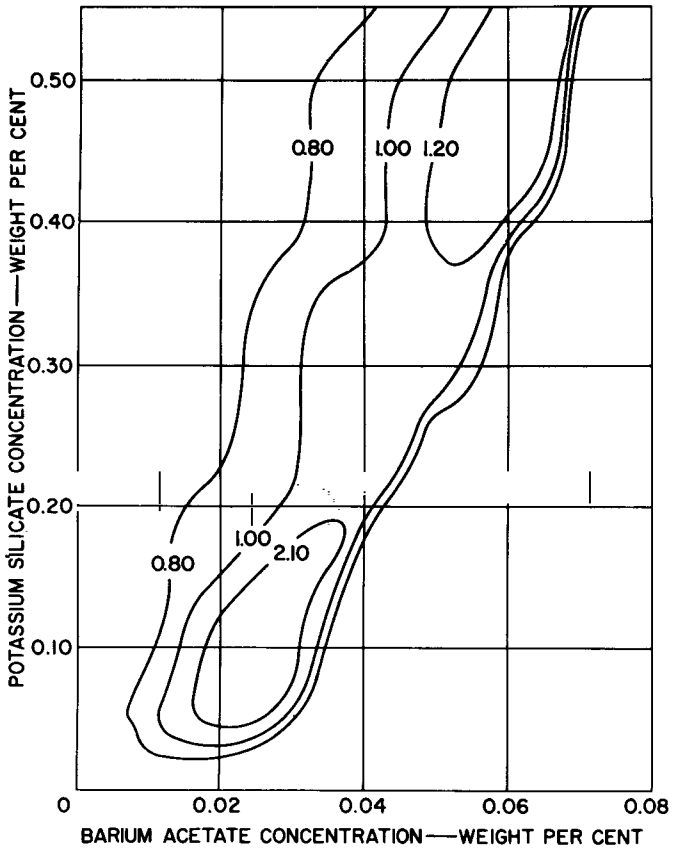


Figure 2. Contour Diagram of Adherence as a Function of Silicate and Acetate Concentration

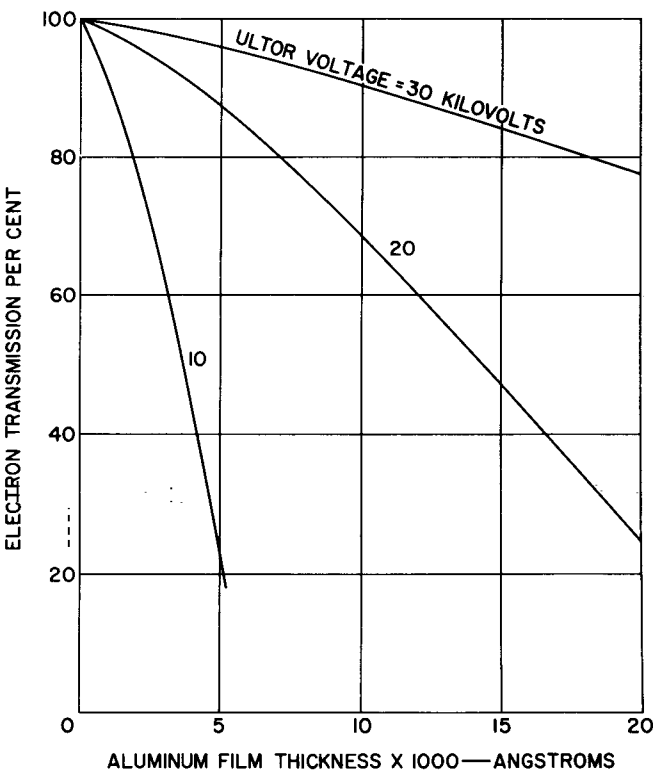


Figure 3. Effective Electron Transmission of Aluminum Films as a Function of Thickness for Various Voltages

Aluminum films are deposited by evaporation at a pressure of 10^{-4} millimeters of mercury. The aluminum is obtained in the form of pellets or wires which are loaded onto, or attached to, tungsten-coil evaporators. The thickness of the evaporated aluminum film is in the range of 1000 to 3000 angstroms.

Evaporating aluminum directly on to a phosphor screen does not result in a bright mirror finish and therefore, gives no gain in useful light output. As a result, it is first necessary to deposit a smooth organic film on the back of the phosphor coating to bridge its rough surface and to support the aluminum film. Since the organic film must later be removed by baking, as little as possible should be used. The general practice is to first wet the phosphor coating with water. The water fills the voids in the screen. Next, the organic film is formed by spraying a film former, such as butyl methacrylate, or by floating a nitrocellulose film, onto the wet coating. After the film has formed and the coating is again dry, aluminum may be evaporated onto the film. The deposited aluminum must have sufficient porosity, usually in the form of pinholes, to allow the decomposition products of the organic layer to escape during screen bake-out.

Some electron energy is lost by the electron beam in passing through the aluminum film. Fig. 3 shows the effective electron transmission for aluminum films at various voltages as a function of film thickness.

SCREENS FOR COLOR TUBES

Phosphor screens for color tubes are entirely different in structure from those used for black-and-white picture tubes. A color-tube screen for a tube such as the 21CYP22 consists of about 1,000,000 phosphor dots, 18 mils in diameter (330,000 each of phosphors for the three colors—red, blue and green). These phosphor dots are positioned accurately with respect to an aperture mask and the three electron guns in the neck of the tube. The geometry of the tube is such that the beam from any gun of a properly operated tube can land only on the phosphor dots for a single color.

The dot array is produced by a photographic system which involves four major steps: application of phosphors, application of a photosensitive binder, exposure to light, and developing.

The phosphors most widely used for color picture tubes are blue-emitting, silver-activated, zinc sulfide; green-emitting, manganese-activated, zinc orthosilicate; and red-emitting, manganese activated, zinc orthophosphate. These phosphors are all white powders and have about one-half the average particle size of water-settling phosphors for black-and-white tubes.

Several different photobinders have been used in the preparation of color-tube screens. These include materials such as: dichromated albumen, dichromated polyvinyl alcohol, Kodak photo resist (KPR), and Kodak photo lacquer (KPL). Dichromated polyvinyl alcohol is a commonly used material. It is available as a water-soluble material which produces a film when spread out as a thin layer and allowed to dry. The dried film may be redissolved in water.

When 2 to 10 per cent by weight of a dichromate salt,

such as ammonium dichromate (in older literature it is often called bichromate), is added to the polyvinyl alcohol (PVA) while it is in solution, the resultant dried film may be made insoluble in water by exposure of the film to ultraviolet radiation. Actually, the photosensitivity extends over quite a wide range as shown in Fig. 4.

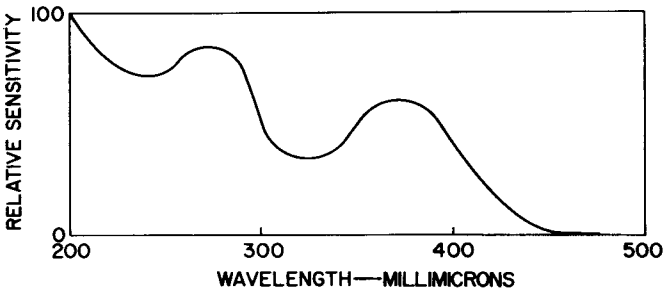


Figure 4. Photosensitivity of PVA-Ammonium Dichromate Film

If only selected portions of the film are exposed to ultraviolet radiation, only those areas become insoluble (other terms are insolubulized, hardened, tanned); the remaining unexposed areas are still water soluble and may be removed by washing them away.

EXPOSURE SYSTEM

A high-pressure, mercury-capillary-type, quartz-envelope lamp is a suitable source for an exposure unit (lighthouse). Its spectrum is typical of that of a high-pressure mercury arc. If a suitable orange filter is placed over the light source, beam landing may be checked with a screened faceplate panel without affecting the photobinder. Although dichromated PVA has very low photosentivity as compared to that of a silver halide photographic emulsion, it must be handled under a safety light which is yellow to orange. White incandescent lamps, fluorescent lamps, or natural daylight are all capable of fogging dichromated PVA films.

DEVELOPING

Water is the developer for dichromated PVA. The water is sprayed onto the screen uniformly with considerable force (20 to 30 pounds per square inch). The water pressure and duration of development are adjusted to give good clean-up, i. e., complete removal of unexposed and unwanted areas with as little harm as possible to the desired screen areas.

The basic elements of photodeposition, phosphor application, photobinder application, exposure, and developing, have been combined in many ways to yield the following systems:

1. One-Step Slurry. Phosphor is mixed with phosphor binder and applied to tube face panel as a slurry.
2. Two-Step Slurry, or Tacky Film. Clear photobinder is applied by spinning. Dried film is exposed and developed. Phosphor slurry in unsensitized PVA is applied by spinning and is then dried and redeveloped.
3. Dusting. Clear photobinder is applied by spraying. While film is still wet, dry phosphor powder is air-settled onto wet film. After the phosphor-photobinder layer is dried, it is exposed and developed.
4. Water Settling. Phosphor powder is applied to panel by conventional water settling. Clear photobinder is applied to dry screen by swirling. After photobinder has soaked into settled screen and dried, the screen is exposed and developed.

All four methods are capable of producing high-quality screens. The one-step slurry method has reached the highest degree of perfection. The two-step slurry method is capable of giving the finest elemental screen area definition (since exposure is made without the presence of light-scattering phosphor crystals). The dusting method can yield screens with the optical characteristics of water-settled screens (high contrast, low optical contact). The water-settling method is the easiest way to make a few screens for experimental purposes.

ONE-STEP SLURRY PROCESS

The principal steps in screening faceplate panels by the slurry process are given in Fig. 5. Each color phosphor is made into a suitable slurry in its own special way. In order to conserve space, only the preparation of the green slurry will be presented as representative of a currently used typical procedure and compositions. The phosphor powder is ball milled in water for one and one-half to three hours. Following the milling, PVA is added from a 10 per cent stock solution to yield a concentration of 38 per cent phosphor, 3.6 per cent PVA, 58 per cent water. This mixture is rolled in the ball mill an additional one and one-half hours to ensure good mixing. After the mill charge is transferred to a portable mixing vat, a final adjustment of slurry viscosity is made by adding more water. A few hours before the slurry is to be used, it is sensitized with ammonium dichromate from a 10 per cent stock solution.

The final slurry composition, as presently delivered to the factory, is 38 per cent phosphor, 3.6 per cent PVA, 6 per cent ammonium dichromate (based on weight of PVA) and 58 per cent water. The slurry is kept in constant agitation until used. It is applied to the faceplate panel by a combination spinning and tilting cycle. The panel is positioned at 15°, open end up. About 180 milliliters of slurry are added to form a puddle near the center of the panel. The puddle is worked out to the edge and spread over the entire faceplate area by spinning and tilting. The wet phosphor slurry film covering the panel is kept in position by additional spinning until it is dry.

The amount of PVA used in the slurry controls the viscosity, which is adjusted to give the best application characteristics. The amount of phosphor in the slurry is essentially the greatest amount that can readily be held in good suspension. This amount, in turn, is controlled by the amount of PVA present. The concentration of ammonium dichromate is not critical. The exposure time decreases with increasing dichromate concentration, but dot quality and slurry shelf-life are poor at very high dichromate concentrations. The effect of dichromate concentration on exposure time is shown in Fig. 6.

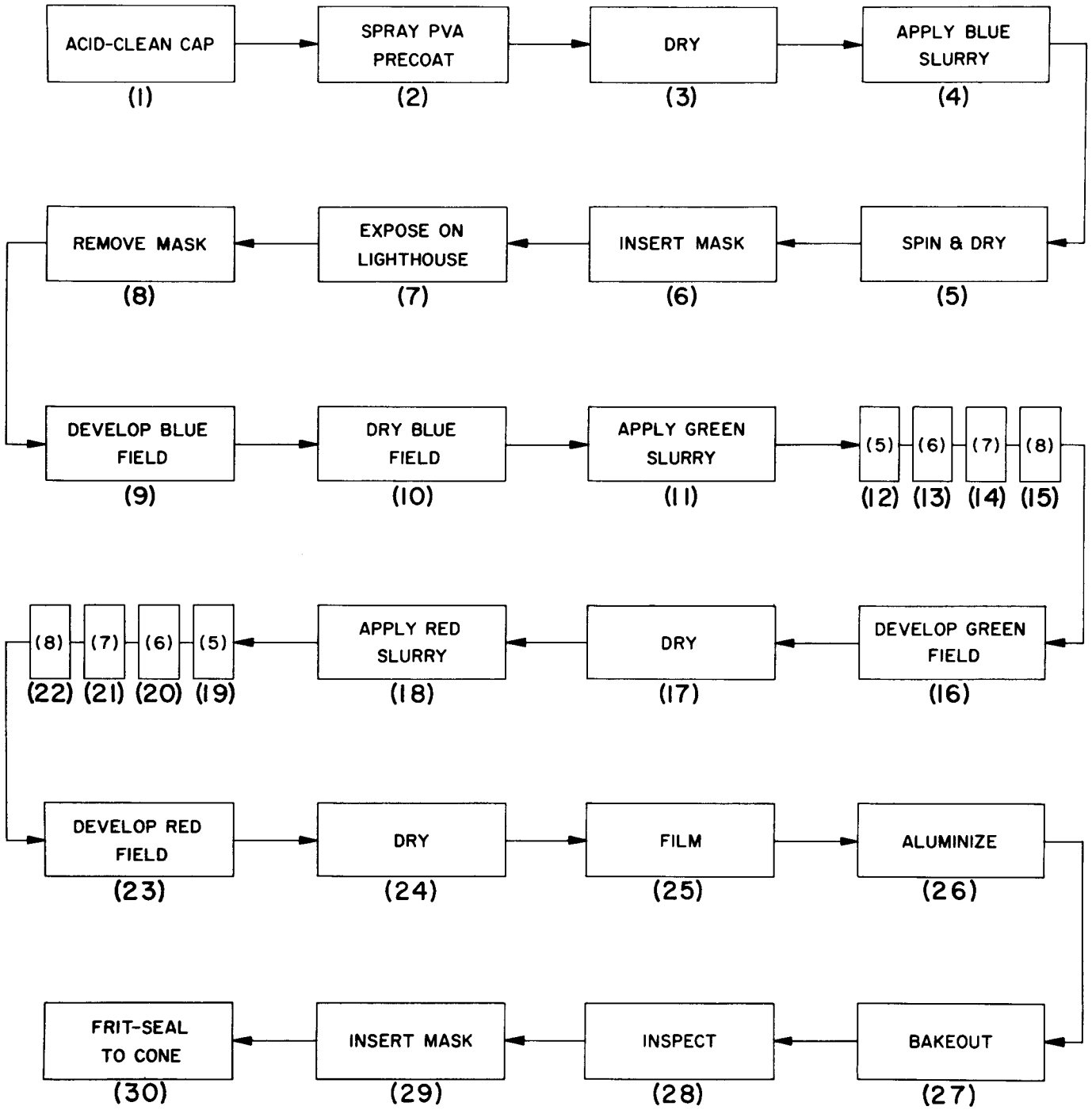


Figure 5. Principal Steps in Screening Face-Plate Panels by the Slurry Process

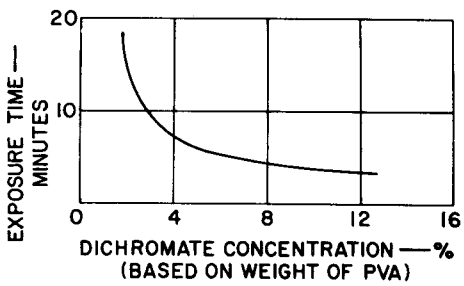


Figure 6. The Effect of Dichromate Concentration on Exposure Time

Phosphor-dot size is dependent, not only on light-source and mask-hole sizes, but also on exposure time. An empirical relationship of exposure time versus hole size is shown in Fig. 7. Final dot size is also dependent to some extent on the pH of the developing water (low pH produces small dots) and the vigor with which the screen is developed. Due to the intensity distribution of light within the light-spot image (see Fig. 8), the core of the phosphor dot is more completely hardened than the edge. In some instances the edge of the dot is not hardened all the way through to the glass substrate and, as a result, the edge, while still wet, is seen to float like a skirt

surrounding the central core. Very hard development will completely break off this skirt, leaving an undersized dot. Prolonged exposure will result in good adherence of this skirt to the glass, giving an oversize dot with poorly defined edges.

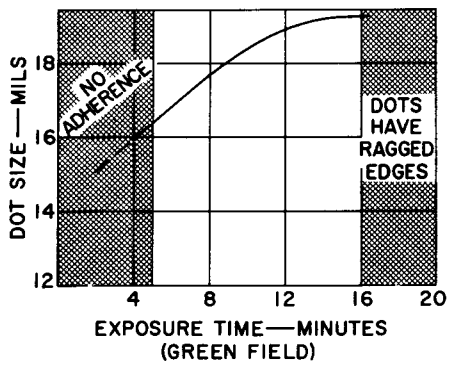


Figure 7. Effect of Exposure Time on Dot Size

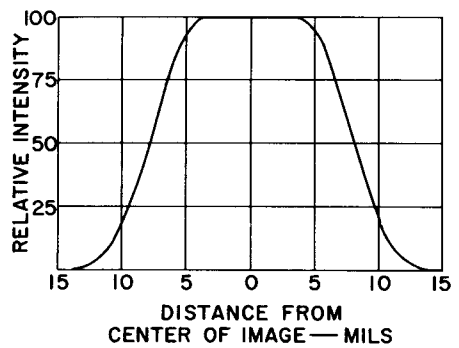


Figure 8. Light Distribution Across an Aperture Image

SCREEN DEFECTS

Cross-Contamination

One of the common problems in making the phosphor screen for a color picture tube is the presence of wrong-color phosphor particles on dots of the desired color. Because phosphor dots are coarse and rough on a microscopic scale, there is a tendency for the phosphors that are deposited on dots already formed, to be trapped mechanically. Insufficient development will result in a high level of this kind of contamination, called "cross-contamination."

In addition to mechanical trapping, some phosphors seem to have a natural affinity for one another. The present sequence of depositing the phosphors, i.e., blue, green, and red, results in the least cross-contamination.

White Nonuniformity

The production of a white field requires that the three color fields be excited simultaneously by their respective guns, and that the beam current for each color be adjusted to give the proper balance for white. If the light output of any field varies over the field, the overall color balance will also vary to produce different whites. Thus, a section of the white field may appear pink relative to the rest of the field.

The causes of white nonuniformity include: (1) mechanical causes (misregistry, dented masks, small dots, overlapping dots, missing dots); (2) electrical causes (stray electron excitation); (3) chemical causes (phosphor deterioration or poisoning); (4) physical causes (variation in screen weight per square centimeter).

Phosphor light output is a function of coating thickness or screen weight per unit area as shown in Fig. 9. As a result, screen weight per square centimeter within a field must be held within a certain range if the light output is to remain essentially constant. Although variations in coating weight per square centimeter in excess of twice this range will affect light output to the extent of only 5 to 10 per cent, this variation may degrade white uniformity.

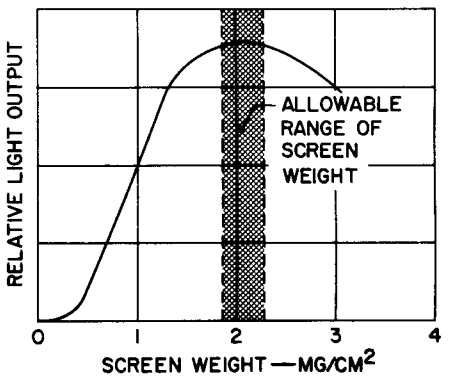


Figure 9. Light Output as a Function of Screen Weight

Blue-Phosphor Breakdown

No discussion of color tube screen defects would be complete without a brief mention of the long-standing problem of blue-phosphor breakdown. The blue phosphor, silver-activated zinc sulfide, is very susceptible to contamination by copper (0.01 parts per million is detectable.

If the blue-screen phosphor becomes contaminated with copper before the screen-baking stage, the baking will cause activation of the zinc sulfide by the copper and a resultant color change from blue to green. Most screens will have a few sections, perhaps only a dot or two, where copper particles have affected the coating. The principal source of copper is the atmosphere. Accordingly, elaborate filtering of the air is required in the screen-processing areas.

FILMING AND ALUMINIZING OF COLOR-TUBE SCREENS

Until very recently color screens were filmed in the same general way as black-and-white screens. However, a new filming technique, known as emulsion filming is now available. The emulsion is similar in principle to latex water-base paints wherein an organic film

former is dispersed in water. After the emulsion is applied by a slurry technique, it is allowed to dry while spinning is continued. As the water evaporates, the organic molecules get closer together until finally they coalesce and form a true organic film. Emulsion filming, particularly for a separate faceplate panel, is cheaper, safer, and more readily automated than conventional spray filming.

The process of aluminizing color screens is similar to that employed for black-and-white screens. Because the evaporator does not have to be inserted through the neck of the color tube, four widely spaced evaporators are used. The use of several evaporators is particularly fortunate because they facilitate obtaining a thick aluminum film. The aluminum thickness in the color tube is in the range of 3000 to 4000 angstroms which is about two times as great as that for black-and-white tubes. This very thick aluminum is needed to absorb low-velocity secondary electrons and to attenuate high-velocity scattered primary electrons, which, if they reached the phosphor screen, would degrade field purity and white uniformity.

SCREEN BAKEOUT

The final step in the processing of color-tube screens is screen baking. The screened, filmed, and aluminized panels are placed on a flat horizontal wire-mesh conveyor belt which carries the panels through a six-zone lehr oven. Fig. 10 is a plot of a typical oven temperature calibration run and shows the temperature in each zone.

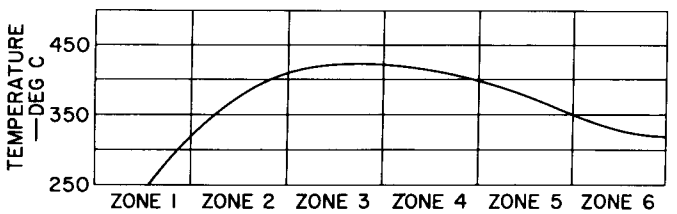


Figure 10. Typical Oven-Temperature Calibration Chart

During the lehr-oven bakeout, the PVA and filming lacquer are removed by decomposition, combustion, evaporation, and sublimation. All of the organic material must be removed if a good vacuum is to be maintained in the finished tube. All of the organic material removed by the baking must pass through the many pinholes normally present in the aluminum film. Filming and aluminizing techniques which do not result in adequate aluminum-film porosity result in blistering of the aluminum.

Recent radiotracer studies give quantitative data, Fig. 11, on the effectiveness of time and temperature in removing the PVA.

PHOTOMETRY OF PHOSPHOR SCREENS

Seeing is such a subjective act that countless systems of light units and definitions have come into existence. It is helpful to go back to the old concept of a single candle placed at the center of a sphere having a radius of one foot. By agreement, let the candle have an intensity of one candle.

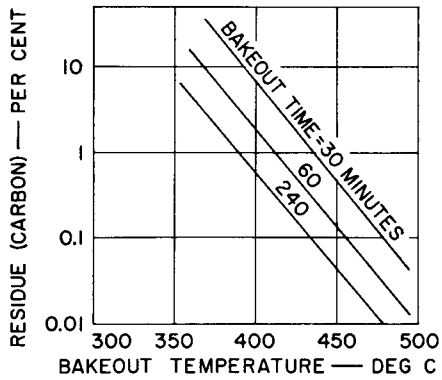


Figure 11. The Effect of Time and Temperature on the Removal of PVA

Further assume that the candle emits uniformly in all directions. It is then convenient to divide space surrounding the candle into unit solid angles with their apices at the candle. The unit solid angle used is the steradian. The candle flame is surrounded by 4π lumens.

A one-square-foot section of the sphere with a radius of one foot subtends a solid angle of exactly one steradian. It follows that the illumination on the one square foot is one lumen. By definition, one lumen per square foot is equivalent to an illumination of one foot candle. If the illuminated surface is fully diffusing without absorption or specular reflection, it will have a brightness of one foot lambert. If the surface absorbs one-half of the incident flux, its brightness will be one-half foot lambert.

An extended light source, such as a cathode-ray-tube screen, has a candle-power equivalent. A phosphor screen is essentially a Lambertian emitter: its candle power intensity varies with the cosine of the angle of view. As a result, a one-candle extended source emits a total of only π lumens on one side.

For example: In Fig. 12, S is a phosphor screen. I_0 is the candle power of the screen, measured on its axis at a distance from S.

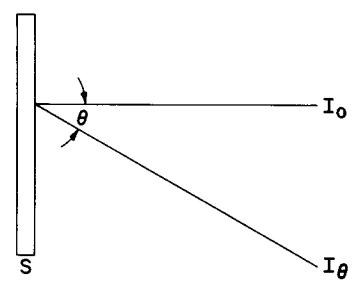


Figure 12. Derivation of Observed Brightness of an Extended Source

Candle power at an angle θ from the axis is:

$$I_0 = I_\theta \cos\theta$$

and the apparent brightness of S is:

$$B = \frac{I}{S}$$

but, $I = I_0 \cos\theta$

and $S = S_0 \cos\theta$

Therefore, $B = \frac{I_0 \cos\theta}{S_0 \cos\theta} = B_0$

and the observed screen brightness is independent of direction of view.

The basic equations for the cathode-ray-tube-screen light source are as follows:

One candlepower (Cp) (extended source) = π lumens (on one side) (1)

One Cp/ft^2 = π lumens/ft^2 = π foot lamberts (2)

The candlepower of a source = $\frac{\text{area (ft}^2\text{) foot lamberts}}{\pi}$ (3)

or, foot lamberts = $\frac{\text{Cp } \pi}{\text{area (ft}^2\text{)}}$ (4)

For example:

A color tube which has a screen area of 250 square inches (1.735 square feet) with a brightness of 8 foot lamberts, has a candle power equivalent of:

$$Cp = \frac{1.735 \times 8}{\pi} = 4.42$$

and emits 4.42 x π = 13. 9 lumens.

If the tube operates at a voltage of 25,000 volts and 400 microamperes (25,000 x 400 x 10^{-6} = 10 watts), the efficiency of the screen itself as a light source under these conditions is 13.9/10 = 1.39 lumens per watt.

MEASURING INSTRUMENTS

The most widely used device for measuring screen brightness is a photovoltaic cell equipped with a correction filter to simulate the color sensitivity of an average eye. This type of instrument is usually calibrated in terms of illumination from an incandescent lamp which has been calibrated by the National Bureau of Standards.

If the photocell unit is used at distances which are large compared to the size of the phosphor screen, readings are obtained in foot candles, and reliable values for screen brightness may be calculated. The general practice, however, is to place the cell on the tube face and, under this condition, brightness is supposedly measured directly. However, the large ac component in the photocell current due to raster scanning and light losses resulting from rays reaching the photocell face at oblique angles can introduce very large errors.

An early subjective instrument used for illumination measurements is the MacBeth illuminometer. This instrument requires matching the brightness of a built-in incandescent lamp with the brightness of the phosphor screen. Because the color of the phosphor light and that of the lamp are quite different in most cases, much experience is required for the proper use of this instrument. Another instrument, known as the Spectra Spot Brightness Meter, uses a lens system and a multiplier phototube. It reads directly in foot lamberts and can measure areas as small as 0.030 inch in diameter. This instrument also presents problems in reading the absolute brightness of a phosphor.

It is recommended that all photometers for measuring color television screens be calibrated with the standard diffuse sources available from the National Bureau of Standards (NBS). NBS is preparing a new diffuse source for measurement of black-and-white screens to replace the original standard which is no longer available.

Measurements made with an eye-corrected photovoltaic cell, Weston Model 759, calibrated on the NBS color primary standards at a distance from the source of three times the source diameter, gave the basic efficiencies shown in Table III.

Table III

Tube Type	Screen Phosphor Number	Phosphor	Efficiency lumens/watt
Display Storage Tube	P20	ZnCdS:Ag	60
21CEP4	P4	ZnS:Ag + ZnCdS:Ag	36
5ABP1	P1	Zn_2Si0_4:Mn	30
21CYP22	P22	Blue—ZnS:Ag Green—Zn_2Si0_4:Mn Red—$Zn_3(P0_4)_2$:Mn	1.39 7.3*
5WP11	P11	ZnS:Ag	27
5WP15	P15	ZnO:(Zn)	18
5ZP16	P16	$CaMgSi0_3$:Ce (ultraviolet)	0.09

*Corrected for mask loss.

COLOR MEASUREMENT AND SPECIFICATION

The basic instrument for measuring color of phosphor screens is a spectroradiometer. This instrument determines the emission intensity of a source as a function of wavelength over a certain wavelength interval. It is often confused with a spectrophotometer which usually can only measure the fraction of incident light that is transmitted or is reflected by a substance. Curves obtained with a spectroradiometer are called spectral-energy-distribution curves (S.E.D. curves).

Because colors are not uniquely determined by S.E.D. curves, such curves have limited direct value. However, starting with the S.E.D. data, it is possible to convert to an objective color specification system such as the C.I.E. (Commission Internationale de L'Eclair-

age) color system-formerly I. C. I. (International Commission on Illumination). In this system, every color is uniquely determined by a set of x, y, and z coordinates. Because x, y, and z are coefficients which add up to one (see below), only two are required to define a color. By convention, x and y are used.

The C.I.E. color system cannot be discussed in detail here; however, the following defining equations are vital to the use of the system:

$$x = \frac{X}{X + Y + Z} \tag{5}$$

$$y = \frac{Y}{X + Y + Z} \tag{6}$$

$$z = \frac{Z}{X + Y + Z} \tag{7}$$

from the preceding equations

$$\frac{X}{x} = X + Y + Z \tag{8a}$$

$$\frac{Y}{y} = X + Y + Z \tag{8b}$$

$$\frac{Z}{z} = X + Y + Z \tag{8c}$$

where x, y, and z are the trichromatic coefficients
X, Y, and Z are the tristimulus values

The conversion from the S.E.D. data to the x and y coordinates is accomplished by a set of three integrations where, in effect, the source is resolved into the relative amounts of the C.I.E. primaries required to color-match the source.

A plot of the x and y coordinates for the pure-spectrum locus produces a horseshoe-shaped locus in Cartesian coordinates. This locus with a straight line joining the ends is called the color mixture diagram. All colors will have coordinates lying within this area.

A very useful characteristic which applies to the mixing of phosphor colors (a color-addition process) is the fact that a straight line joining two phosphor points represents all the intermediate colors that can be obtained from mixtures of these two phosphors. For example, it can be seen in Fig. 13 that combining a blue phosphor and a yellow phosphor in the proper proportions will produce white. It is this principle that is used in the making of screens for black-and-white picture tubes. Color mixture problems may be solved by considering them as moment arms. The two colors have a certain intrinsic colormetric weight, called stimulability, which must balance around the mixture point as a fulcrum. Stimulability is defined as luminous efficiency Y (lumens per watt) divided by y. $S = Y/y$

It has been shown that Y/y is equal to $X + Y + Z$ which are the actual amounts of the C.I.E. primaries required to match the unknown source in color and intensity.

APPLICATION OF C.I.E. COLOR SYSTEM TO BLACK-AND-WHITE SCREENS

The development of an all-sulfide black-and-white screen requires the making of a series of silver-activated zinc-cadmium sulfide phosphors. The series should run from 100 per cent zinc to 30 per cent zinc (the balance cadmium sulfide) in steps of approximately 10 per cent. The color of each phosphor is measured and plotted on the C.I.E. mixture diagram. A line through these points is the color locus of the zinc-cadmium sulfide system and is shown in Fig. 14. In addition to color, the visual efficiency must be measured— at least on a relative scale—and the stimulabilities calculated. Plots of relative visual efficiency and relative stimulability are shown in Figs. 15 and 16.

Neglecting certain secondary effects, such as the body color of the yellow phosphor, it is now possible to pick a pair of phosphors which produce a given shade of white with maximum light output. In referring to white light, it is convenient to describe it in terms of the nearest equivalent black-body temperature. Charts are available for the center section of the C.I.E. diagram which show the black-body color locus, isothermal lines across this screen, and a minimum perceptible color difference scale (M.P.C.D.) along the isothermal lines.

Assuming the 10,000 degrees Kelvin at +20 M.P.C.D.'s (above the black-body line) is a desirable white, it can be seen from Fig. 17 that a number of straight lines can be drawn through this point, each of which will intersect the phosphor color locus at two places.

The choice of the phosphor pair depends simply on which pair produces the greatest light output.

The screen brightness Y_{sc} of a television tube is related to the intrinsic brightness of the phosphors by the relationship:

$$Y_{sc} = \alpha Y_y + \beta Y_b \tag{9}$$

where Y_y and Y_b are the luminosities of the yellow and blue phosphors, respectively, and α and β are the relative amounts used in the blend. Since Y_y is usually about ten times as great as Y and α and β are nearly equal, most of the visual efficiency of a television screen is due to the yellow phosphor.

The approximate amounts of blue and yellow phosphor needed to make the 10,000-degree Kelvin white light are given by the following equation:

$$\alpha = \frac{(100 - \alpha)\, S_B \cdot L_B}{S_y \cdot L_y} \tag{10}$$

where,

α = per cent yellow to be used
S_y = relative stimulability of the yellow phosphor
L_y = linear distance from the 10,000-degree Kelvin point to the yellow end of the color line
S_B = relative stimulability of the blue phosphor
L_B = linear distance from the 10,000-degree Kelvin point to the blue end of the color line

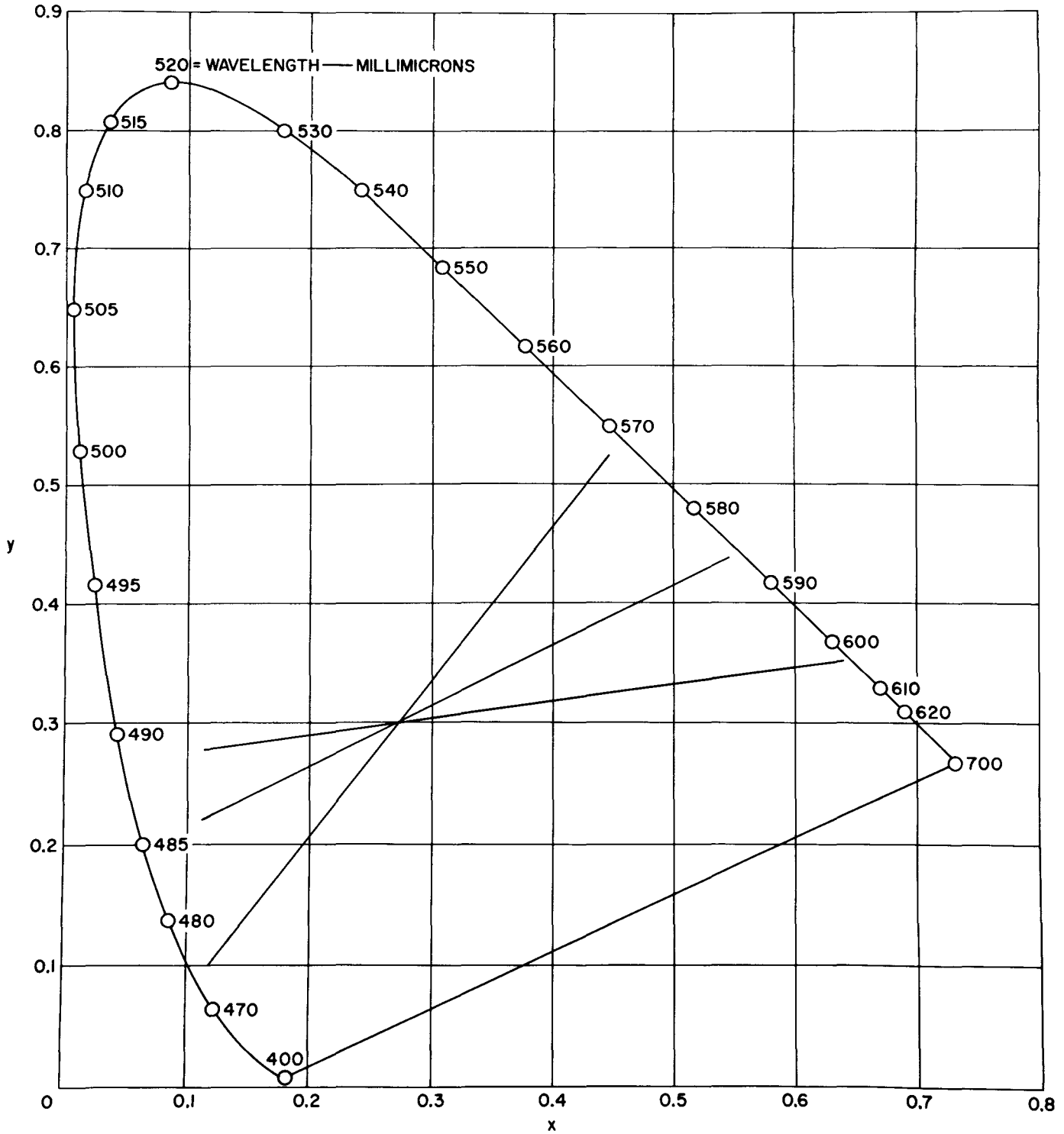


Figure 13. Three Complementary Phosphor Pairs which will Produce White

A pair of phosphors for optimum efficiency may be picked from among the three pairs shown in Fig. 17 by means of Eqs. (9) and (10).

Table IV shows that phosphor pair number one will produce a brighter 10,000-degree Kelvin white than any of the others.

Table IV

Phosphor Pair	Per Cent Yellow to Make $10{,}000^{\circ}$ Kelvin White	Relative Screen Brightness
1	61.80	100.00
2	30.53	57.90
3	26.27	41.60

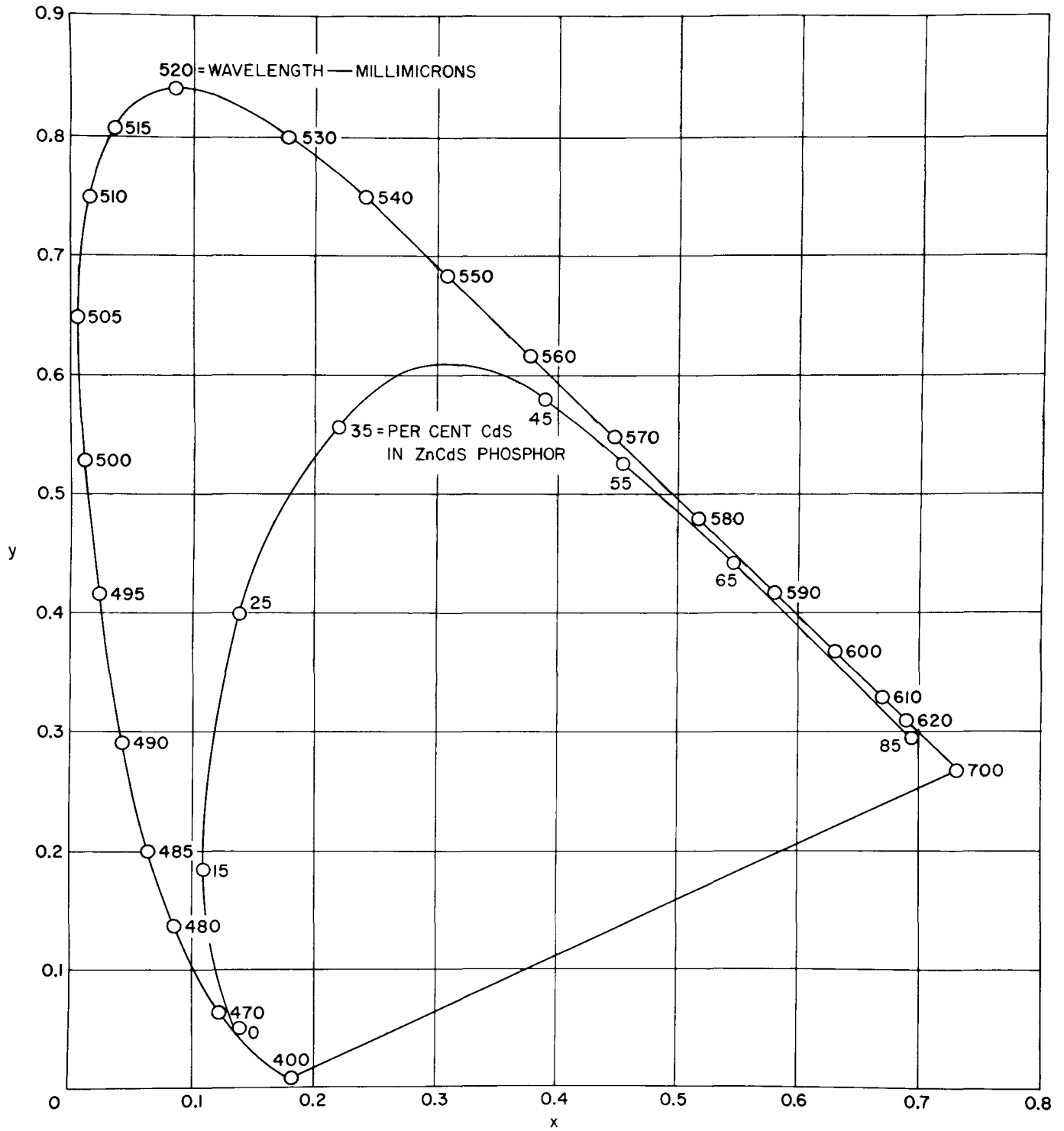


Figure 14. Phosphor Locus of the Zinc-Cadmium Sulfide Silver-Activated System

Calculating a 10,000-degree Kelvin white on paper is simple and direct. Achieving it in a finished tube is very difficult. Factors such as body color, screen weight, particle size, and operating conditions all affect the final result.

APPLICATION OF C.I.E. COLOR SYSTEM TO COLOR TELEVISION SCREENS

Much of what has been said about photometry, colorimetry, and the use of the C. I. E. system for screens for black-and-white tubes also applies to color screens for tubes. Early in the development of color television the National Television Standard Committee (NTSC) proposed the three phosphor primaries and described them in terms of their C. I. E. coordinates. (See Table V.)

The C. I. E. diagram has been used often to show how the fidelity of color television pictures compares with

other color media such as color photography and color printing. Fig. 18 compares the color gamut of the color television phosphors with the color gamut of printing inks.

Table V

Color	NTSC PROPOSAL C.I.E. Coordinates	
	X	Y
Red	0.670	0.330
Green	0.210	0.710
Blue	0.140	0.080

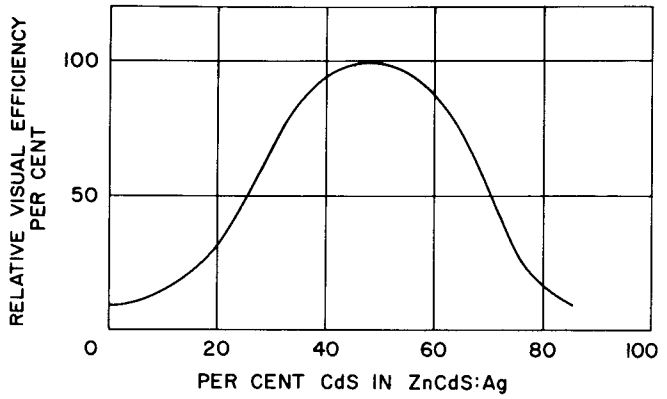


Figure 15. Relative Efficiency of a Phosphor

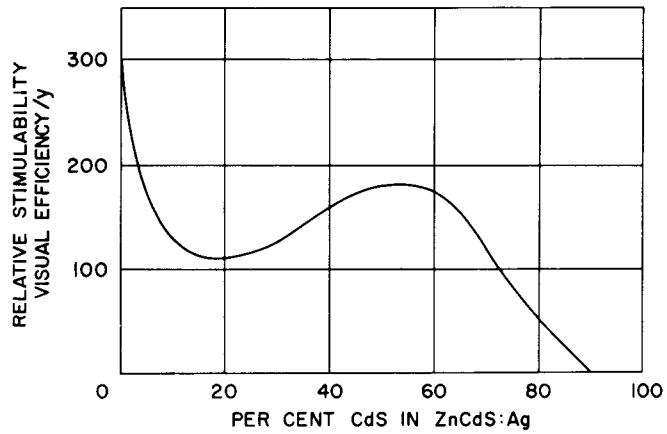


Figure 16. Relative Stimulability of a Phosphor

The laws of color addition may be used to determine the ratios of beam current for the red, blue, and green guns to produce any color, including white, within the phosphor triangle. These same principles may be used to determine the magnitude of color change when changes are made in the beam-current ratios for the three guns.

It is also possible to calculate the relative and absolute currents for each gun to produce any color point within the color triangle by means of Grassmann's equations. Simply stated, these equations say that the X, Y, and Z, which are the C.I.E. tristimulus values of a color mixture, are equal to the sum of X, Y, and Z of the component colors.

Applying these equations to the problem of finding the gun currents which will yield 8500-degree Kelvin white gives:

$$X_{8500} = I_R X_R + I_B X_B + I_G X_G \quad (11)$$

$$Y_{8500} = I_R Y_R + I_B Y_B + I_G Y_G \quad (12)$$

$$Z_{8500} = I_R Z_R + I_B Z_B + I_G Z_G \quad (13)$$

where I_R, I_B, I_G are the relative red, blue, and green gun currents

and, X_R, Y_R, Z_R
Y_B, Y_B, Z_B
X_G, Y_G, Z_G

are the C.I.E. tristimulus values of the red, blue, and green phosphors.

and, X_{8500}, Y_{8500}, and Z_{8500} are the tristimulus values for the 8500-degree Kelvin white point.

The following data for the three phosphors are sufficient to make the calculation as outlined:

	x*	y*	z*	X‡	Y†	Z‡
Red	0.670	0.330	0.000	4	2	0
Blue	0.140	0.080	0.780	3.5	2	20
Green	0.210	0.710	0.080	2	8	1
8500° Kelvin White	0.287	0.316	0.397	-	-	-

*Value calculated from the spectral-energy distribution curves.
†Value measured with eye-corrected photocell.
‡Value calculated from Eqs. (5), (6), and (7).

Using these data:

$$x = 0.287 = \frac{4I_R + 3.5I_B + 2I_G}{6I_R + 25.5I_B + 11I_G}$$

$$y = 0.316 = \frac{2I_R + 2I_B + 8I_G}{6I_R + 25.5I_B + 11I_G}$$

$$z = 0.397 = \frac{20I_B + I_G}{6I_R + 25.5I_B + 11I_G}$$

Solving these equations gives the following current relationships:

	Per Cent Calculated	Per Cent Measured
I_G =	28	30
I_B =	22	20
I_R =	50	50

The final step is to calculate the absolute currents for a given light level.

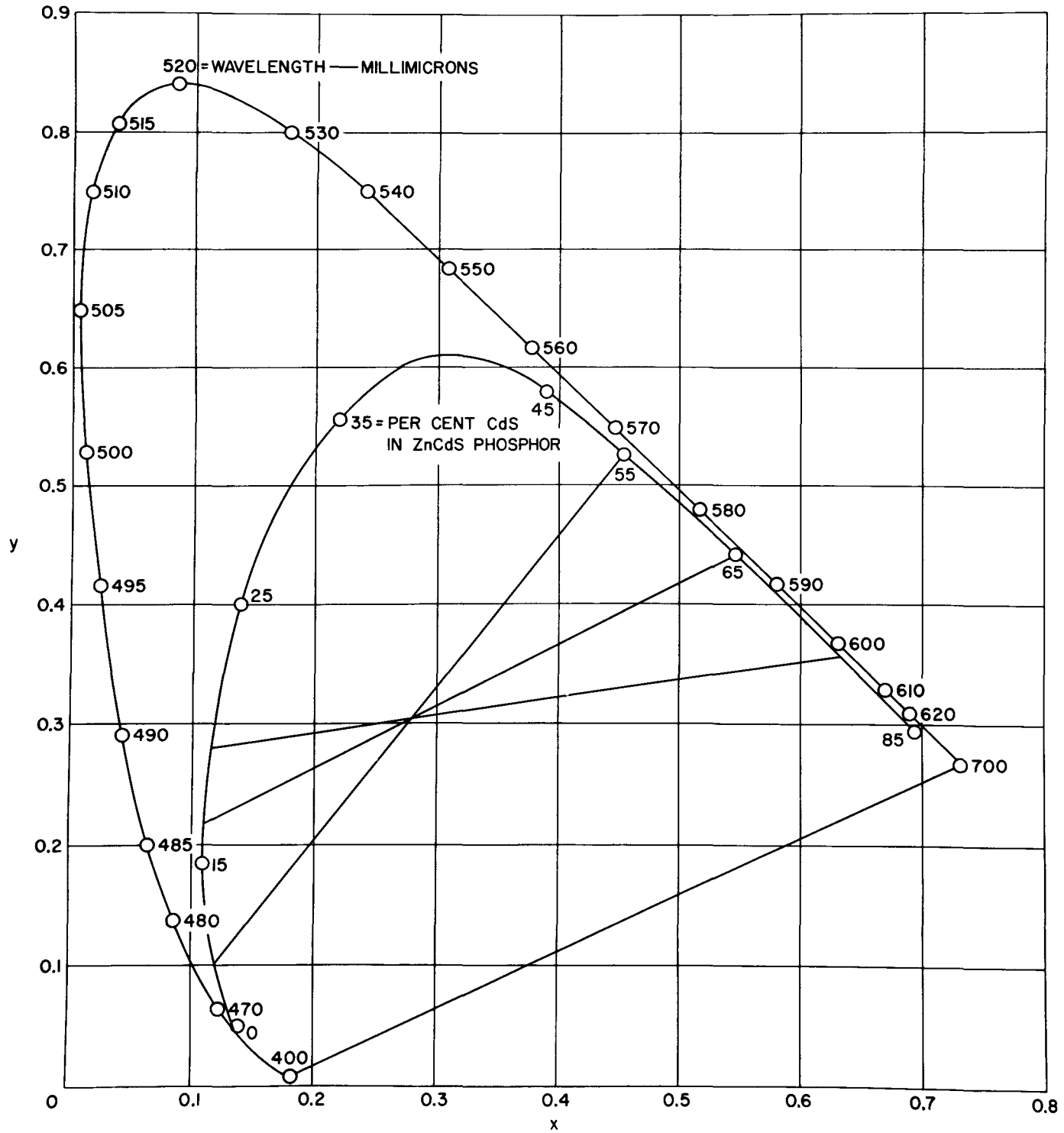


Figure 17. Locus of Zinc-Cadmium Sulfide Phosphor System and Three Complementary Phosphor Pairs for Producing White Light

Assume:

8 footlamberts brightness required		
Screen area		= 250 sq. in.
Ultor voltage		= 25,000 volts
Candle Power		= 4.4 [See Eq. (3)]
Y_{8500}		= 3.7 cp/w [See Eq. (12)]
Beam Watts	$\frac{4.4}{3.7}$	= 1.2
and		
Beam Current		$= \frac{4.4 \times 10^{-6}}{25000}$ 480 μ a

The principles of photometry and colorimetry are relatively simple and straightforward. However, the practical difficulties of making precise measurements are enormous and have yet to be fully resolved. An accuracy of ±10 per cent on measurement of absolute screen brightness or gun current to produce a given brightness is quite good.

An accuracy of ±0.005 in x- and y- C.I.E. units would also be quite good; even this level is reached with considerable difficulty.

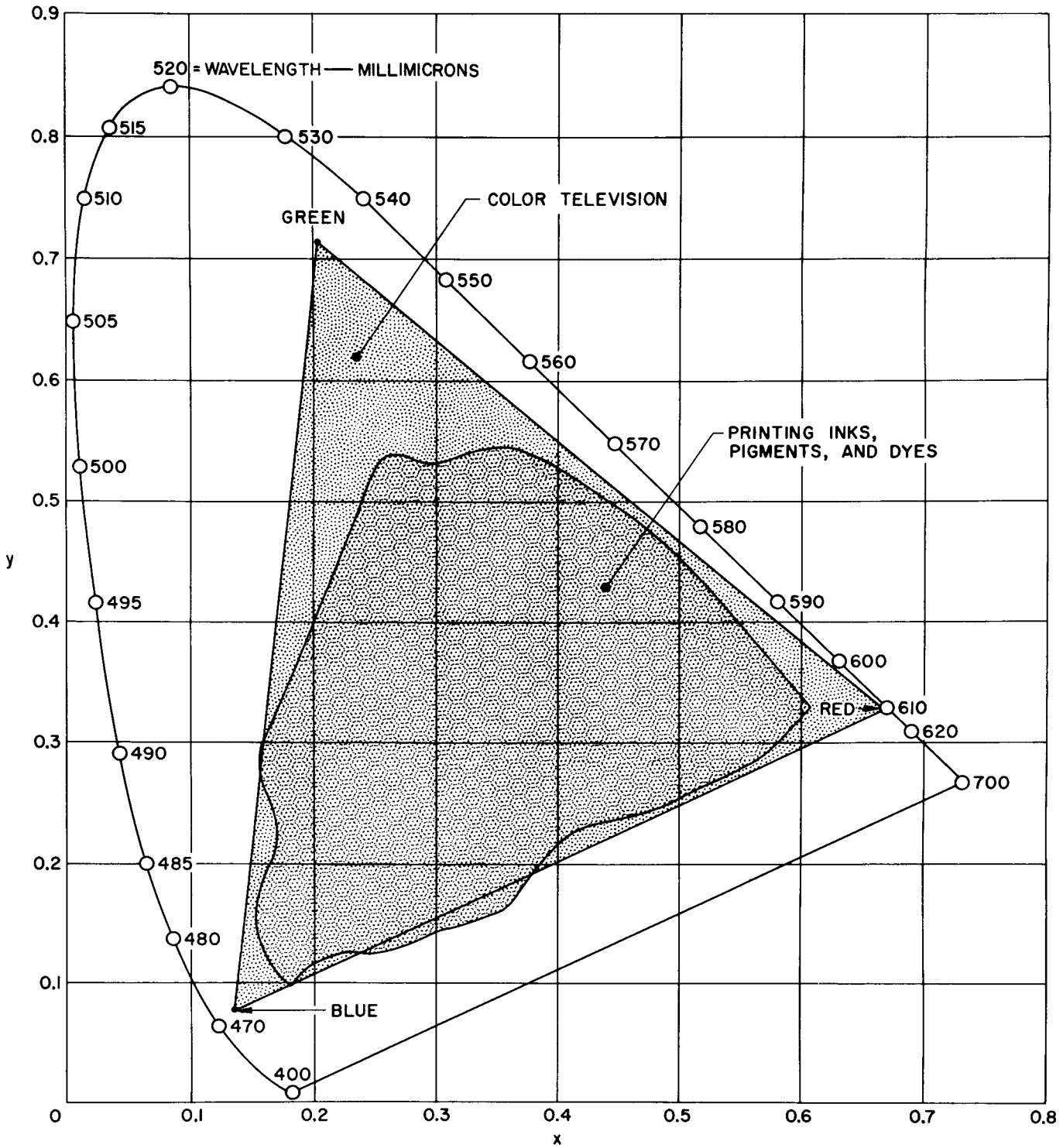


Figure 18. Color Gamut of Color-Television Phosphors Compared with Pigments, Dyes, and Inks

REFERENCES

1. Hardy, A. E., Handbook of Colorimetry, Technology Press, Cambridge, 1936

2. Nottingham, W. B., "Notes on Photometry, Colorimetry and Centibel Scale," Radiation Laboratory Report 804, M.I.T., 1945

3. Judd, D. B., "Equivalent Color Temperatures," J.O.S.A, 26, pp 421-426, 1936

4. Leverenz, H. W., Luminescence of Solids, Wiley, New York, 1950

5. Judd, D. B., Color in Business, Science, and Industry, Wiley, New York, 1952

6. Gibson, K. C., "Spectrophotometry," National Bureau of Standards, Circular 484, September 15, 1949

7. Hardy, A. E., "A Combination Phosphorometer and Spectroradiometer for Luminescent Materials", Jour. Electrochem. Soc., Vol 91, pp 127-146, 1947

8. Hardy, A. E., "Application of I. C. J. Color System to Development of All-Sulfide White Television Screen," RCA Review, Vol. VIII, No. 3, September 1947

9. Epstein, D. W., "Photometry in Television Engineering," Electronics, pp 1-5, July, 1948

10. "Optical Characteristic of Cathode Ray Tube Screens," Report No. J6-C3-1, October 1, 1959 (Available from E. I. A. Standards Laboratory, 32 Green Street, Newark, New Jersey)

Manufacture of Cathodoluminescent Materials

J. A. Markoski

Lancaster

SULFIDE PHOSPHORS

Until the second World War, luminescent materials of the sulfide type were made in small amounts by laboratory methods. The war created a heavy demand for radar phosphors and laboratory production methods proved utterly inadequate. As a result, RCA was asked by the Navy to build the first commercial plant in the United States for the production of radar phosphors.

The process chosen for the new plant was the so-called "acid process" in which zinc sulfide or a mixture of zinc and cadmium sulfides is precipitated from an acid solution of the corresponding salts. The Pyrex glass apparatus used in the laboratory was replaced by glass-lined steel reaction vessels of commercial design. In addition, many special features were provided to meet the very high purity standards necessary for the manufacture of satisfactory phosphors.

Most of the radar screens manufactured in the Lancaster plant during the war were of the cascade (or layer) type in which the layer next to the glass was phosphorescent zinc-cadmium sulfide. This phosphor was excited by absorbing the light from a second layer of blue-emitting zinc sulfide which was placed directly over it. After the war, the composition and the characteristics of these two phosphors were altered to make them suitable for black-and-white television.

The reaction by which zinc sulfide is made is:

$$ZnSO_4 + H_2S \rightarrow ZnS + H_2SO_4 \qquad (1)$$

The by-product, sulfuric acid, is neutralized with zinc oxide and the resulting zinc sulfate is used to make the next batch.

$$H_2SO_4 + ZnO \rightarrow ZnSO_4 + H_2O \qquad (2)$$

Although the basic chemistry appears simple, the actual production of high-grade phosphors is not. Unless many precautions are taken and rigid processing techniques are observed, unsatisfactory phosphors result. Even though some deviations may appear to be trifling or insignificant, they often have harmful effects which persistently defy explanation. In order that some of these pitfalls may be avoided, the essential features of each step will be taken up in sequence.

Preparation of Stock Solutions

Although other soluble zinc salts can be used, the sulfate is best suited for the method of manufacture described here. The greater part of the zinc sulfate is obtained by adding zinc oxide to the by-product, sulfuric acid, obtained from the reaction shown in Eq. (1). Some handling losses in the recovery of the acid are inevitable; these losses are compensated for by the addition of concentrated Chemically Pure (C. P.) grade sulfuric acid. United Stated Pharmacopoeia (U. S. P.) grade zinc oxide is added slowly with continuous stirring until a methyl red indicator shows the alkaline color. At this point, there is no free sulfuric acid, and the pH of the solution is 5.5 to 5.9 at 25 C. This condition is important because it assures an excess of zinc oxide essential for the removal of iron by the reaction:

$$2\,Fe^{+++} + 3\,ZnO + 3\,H_2O \rightarrow 2\,Fe(OH)_3 \downarrow + 3\,Zn^{++} \qquad (3)$$

At one time ammonium hydroxide was added to raise the pH and precipitate iron, but it has since been found that zinc oxide is superior. The addition of an oxidizing agent such as hydrogen peroxide to oxidize any ferrous iron which may be present has been found unnecessary. Since the by-product, sulfuric acid, is nearly saturated with hydrogen sulfide, the addition of hydrogen sulfide is also unnecessary. A partial precipitation of zinc sulfide is automatically obtained and carries down all other less soluble sulfides. After the zinc sulfate solution has been allowed to cool and settle, the mixed sludge of zinc sulfide and excess zinc oxide is removed either by filtration or by siphoning off the clear supernatant solution. It is extremely important to avoid metallic and organic contamination; all reactors, pipes and valves used in the neutralization step must be made of materials impervious to 1.5 to 2.0 normal (N) sulfuric acid.

Cadmium sulfate stock solution is prepared by dissolving Analytical Reagent grade cadmium sulfate in demineralized water and purifying the solution by two partial precipitations. The first precipitate is cadmium hydroxide which carries down all other less soluble hydroxides. This precipitation is performed by adding ammonium hydroxide until the pH is 5.8 to 6.0.

The second partial precipitation is performed by adding hydrogen sulfide, which precipitates cadmium sulfide and all other less soluble sulfides.

Cadmium sulfate solution may also be prepared by the reaction between cadmium oxide and sulfuric acid. This method is cheaper and will probably be adopted in the future. In this case, it is probable that an excess of cadmium oxide will act in a manner analogous to that of zinc oxide in the removal of Fe^{+++} [Eq. (3)].

Precipitation

The principal problems encountered in precipitation are safety, purity, and corrosion.

Hydrogen sulfide is so toxic that no leaks in the system are permissible. Very slow seepage from the stuffing box of the precipitation reactor is practically unavoidable but is not hazardous when proper venting is provided. Other leaks may be prevented by the use of pure gum-rubber gaskets 1/4-inch thick. For seals against liquids, 1/4-inch Neoprene gaskets are the most satisfactory.

Purity and corrosion are interrelated because corrosion is capable of introducing harmful metallic impurities. It has been reported that the adverse effects of the principal contaminants such as iron and copper are detectable in concentrations as low as 1 part in 10 million.

Corrosion is controlled mainly by the use of glass-coated steel. Supplementary construction materials include rubber-lined steel, porcelain, polyvinyl chloride, Teflon, and Hastelloy C. Theoretically, such parts as drive shafts and bearings are not exposed to corrosive atmospheres; nevertheless, experience has shown that gradual erosion takes place and ultimately causes severe damage. Not all contamination is metallic. Grease and oil or shreds of packing material from stuffing boxes can be disastrous to quality. All three can be eliminated by the use of mechanical rotating seals made of graphite and Hastelloy C. The advantages of mechanical rotating seals are offset by high initial cost and complexity. The complicated structure of the rotating seal makes it difficult to repair. Stuffing-box seals are relatively inexpensive and simple to maintain. The best packing material for this type of seal is paraffin-impregnated flax. As the packing material wears, the shreds will fall into the reaction tank; this can be prevented by placing a Lucite shaft cup on the agitator stem inside the tank.

The entire precipitation reaction takes place on the acidic side with the acidity progressively rising as the reaction proceeds. As the acidity increases, the rate of reaction decreases; therefore, strong agitation is needed to save time and to assure 98 per cent complete precipitation. Best results are obtained with a center-agitated reactor equipped with a three-bladed impeller and a baffle with two projecting side arms. The second arm on the baffle is placed at the surface of the liquid to provide the swirling and tumbling which promote gas absorption. A dip pipe made of Hastelloy C will also provide additional effectiveness. Hydrogen sulfide is delivered to the reactor through a pipeline filter containing a disposable filter cartridge. No other purification is necessary.

For the sake of economy, the rate of reaction must be established for each reactor of a given design. The rate was first determined for a 100-gallon reactor of the type described with no dip pipe. The zinc sulfate solution was a 1.6 N and the pH was 5.9. The air in the reactor was swept out with hydrogen sulfide until an initial absorption of two pounds per hour was assured. This rate of flow was maintained until it was no longer possible to precipitate any more zinc sulfide. Atmospheric pressure was maintained in the reactor by allowing the excess hydrogen sulfide to escape to a fume scrubber. Samples of solution were removed at intervals and analyzed for unprecipitated zinc. The results were used to calculate the data required to plot the curve in Fig. 1, which shows the total amount of zinc sulfide precipitated at any given time. The slope of this curve gives the rate of precipitation at any instant provided the proper scale correction is applied. By calculation, pounds of zinc sulfide precipitated per hour can be converted into equivalent pounds of hydrogen sulfide absorbed per hour. When these results are plotted as a function of time, as in Fig. 2, a curve is obtained which shows what the rate of flow of hydrogen sulfide should be at any time during the precipitation.

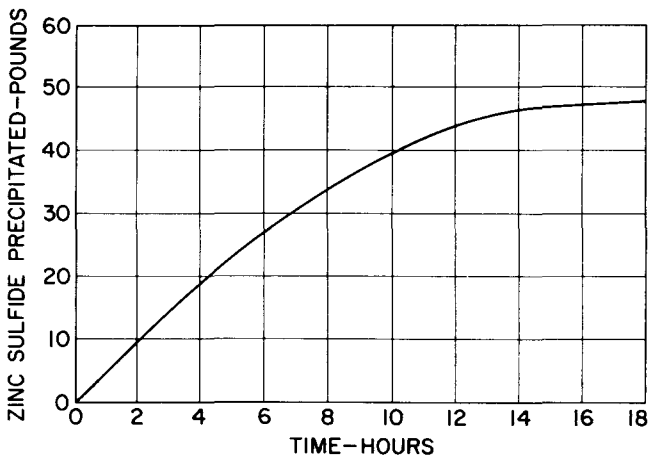


Figure 1. Rate of Precipitation of Zinc Sulfide

Physical properties of the zinc sulfide precipitate are controlled by the rate of precipitation and by the hydrogen ion concentration. Zinc sulfide that has been precipitated very rapidly from an alkaline solution is extremely fine and difficult to wash. When this material dries, it forms hard of glassy lumps which must be crushed and ground. On the other hand, if zinc sulfide is precipitated slowly from an acid solution, a coarser product is obtained which can be washed easily, and while drying, forms a dense free-flowing powder.

Ninety-eight per cent of the zinc can be precipitated from 1.0 to 2.0 N zinc sulfate solution if the initial pH is 5.9 and agitation is continued until no more hydrogen sulfide is absorbed. If it is assumed that the solubility product for zinc sulfide and the ionization constants for hydrogen sulfide are correct, it is impossible to show theoretically that such high yields are obtainable. Therefore, it is evident that the simple equation $Zn^{++} + S^{=} \rightarrow ZnS$ does not adequately describe the precipitation of zinc sulfide.

The ionization constants K_1 and K_2 for aqueous solu-

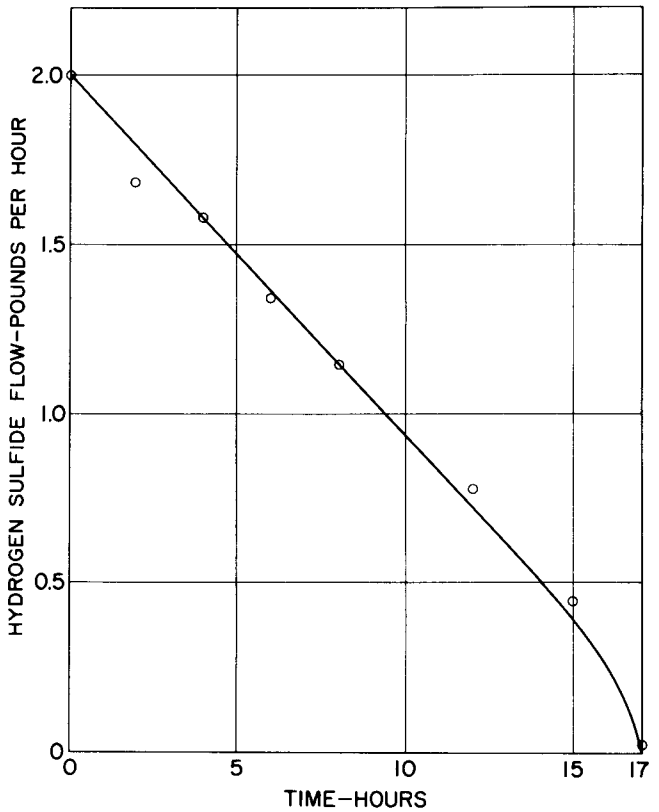


Figure 2. Rate of Hydrogen Sulfide Delivery for Precipitation of Zinc Sulfide

tions of hydrogen sulfide are:

$$\frac{C_{H^+} \times C_{HS^-}}{H_2S} = K_1 = 9.1 \times 10^{-8} \quad (4)$$

and

$$\frac{C_{H^+} \times C_{S^=}}{C_{HS^-}} = K_2 = 1.2 \times 10^{-15} \quad (4a)$$

At 25 C and atmospheric pressure, a saturated solution of H_2S is approximately 0.1 molar (M). By using this concentration and the values given above for K_1 and K_2, it can be calculated that $C_{HS^-} = 9.5 \times 10^{-5}$ and $C_{S^=} = 1.2 \times 10^{-15}$. Thus, it can be seen that the concentration of the HS^- ions is more than 10^{10} times greater than that of the $S^=$ ions. In strong acid solutions, the common-ion effect will decrease the concentration of the $S^=$ ions to an extremely low level, and though the HS^- ions will also be repressed, their concentration will still be very much greater than that of the $S^=$ ions. This suggests the possibility that zinc sulfide is not entirely precipitated by the reaction:

$$Zn^{++} + S^= \longrightarrow ZnS \quad (5)$$

but the two-stage reaction

$$Zn^{++} + 2\ HS^- \longrightarrow Zn(HS)_2 \quad (5a)$$

$$Zn(HS)_2 \longrightarrow ZnS + H_2S \quad (5b)$$

also takes place near the end of the precipitation. The existence of $Zn(HS)_2$ is controversial, but there is good evidence to show that it is present in zinc sulfide to the extent of approximately two per cent and decomposes sharply at 390 C. This subject is discussed in greater detail under "Neutralization and Washing."

If it is desired to shift the emission color from the blue emission color of silver-activated zinc sulfide to some longer wave length, zinc sulfide is precipitated almost completely to build up the acidity, then cadmium sulfate is added and the precipitation is continued until absorption of hydrogen sulfide ceases. Precipitation from a mixture of zinc sulfate and cadmium sulfate should be avoided because this method produces a precipitate which is difficult to wash and is too light and fluffy after drying. Such material is difficult to pack into crucibles and greatly reduces firing capacity.

The shift of emission color is precisely governed by the zinc-cadmium ratio, therefore accurate control of all concentrations, pH's, and yields is essential.

Neutralization and Washing

After the zinc or zinc-cadmium precipitation has been completed, the precipitate is allowed to settle out and the supernatant sulfuric acid is removed and neutralized with zinc oxide. This step constitutes the preparation of zinc sulfate stock solution.

The settling time of the precipitate depends upon the particle size; two or three hours is usually sufficient. The by-product, sulfuric acid, should be removed; washing of the precipitate should be started as soon as possible to minimize the reverse reaction. In the presence of 1 N sulfuric acid, the zinc sulfide not only begins to redissolve, but also forms lumps which eventually cement together into one large mass that cannot be broken up by ordinary stirring. Zinc-cadmium sulfide is not subject to this lumping tendency, but may change in composition because of the tendency of the more soluble component (ZnS) to redissolve.

Sulfide precipitates are washed until the wash water has a pH of 2.5 to 3.0. The remaining acidity is then neutralized by adding ammonium hydroxide. It has been found that this step is indispensable in making ZnS:CdS:Cu phosphors for the P-7 type of radar screen. If omitted, very poor build-up characteristics will result. In the case of the ZnS:Ag phosphor, G. E. Crosby has found that neutralization of the small amount of residual acidity prevents the sudden and troublesome evolution of large volumes of hydrogen sulfide during the subsequent firing operation. It is important to note that lower temperatures are ineffective and hydrogen sulfide is not evolved gradually as would be expected if it were merely absorbed on the surface of the zinc sulfide. These facts indicate the existence of $Zn(HS)_2$ (zinc hydrosulfide) or perhaps an addition compound such as $ZnS.H_2S$. The amount present is determined by drying 10-gram samples of zinc sulfide at 175 C for 48 hours, then baking for 2 hours at 450 C. The loss of weight at 450 C is 0.5 per cent and is due to the evolution of hydrogen sulfide, presumably from the decomposition of zinc hydrosulfide. One gram of hydrogen sulfide is equivalent to 3.86 grams of $Zn(HS)_2$; therefore, the

amount of $Zn(HS)_2$ in this case is slightly more than 1.9 per cent. This phenomenon of gas evolution at 390 C is unnoticed in batches of one kilogram or less, but it becomes a major problem when the operation is scaled up.

From a manufacturing point of view, sodium bicarbonate is superior to ammonium hydroxide for neutralizing residual amounts of sulfuric acid in washed precipitates because it removes all traces of hydrogen sulfide which is both obnoxious and hazardous. This substitution has not yet been adopted because there is a possibility that it may have an adverse effect on the application properties of the finished phosphor.

Although washing by decantation is simple and effective, it has the disadvantage of being excessively time consuming. Essentially the same result has been obtained experimentally in a much shorter time by the use of a rubber-lined centrifuge equipped with an Orlon filter bag. This method is satisfactory except for the fact that hydrogen sulfide is practically impossible to remove unless sodium bicarbonate is used for neutralization of the filter cake.

Drying

There are many ways in which unfired zinc sulfide can be dried. It is not likely that any particular method has a significant effect on quality, as long as care is taken to prevent metallic contamination. In the ordinary batch-drying process, wet zinc sulfide or zinc-cadmium sulfide is placed in 10-liter Pyrex evaporating dishes covered with aluminum lids. These dishes are placed in a hot-air drying oven and heated at 175 C until all water has been removed from the batch.

Activation and Fluxing

Activators may be added either before or after a sulfide is precipitated. In the first case, a soluble salt containing the activator is added to the solution of zinc sulfate and the activator is then co-precipitated with the base material. In the second case, the activator solution is added to a slurry of zinc or zinc-cadmium sulfide which has been previously washed and neutralized.

The subject of fluxes is complex and is treated thoroughly in the literature. For this reason, only those techniques which are specific to the manufacture of cathode-ray tube phosphors will be discussed.

In the preparation of radar phosphors or those which are to be fired above 1000 C, it is extremely important to use very pure fluxes. Sufficient purity can be obtained by dissolving Reagent Grade salts such as sodium, ammonium, or calcium chloride in pure water, and then adding ammonium hydroxide and unfired zinc sulfide. The ammonium hydroxide serves to remove those impurities which have insoluble hydroxides, and the zinc sulfide removes impurities which have sulfides with solubility products below 1.2×10^{-23}.

Fluxes and activators can be added at the same time to slurries of base material, or a system of "dry" fluxing and activation can be used. In the latter case, all of the activator is added to a water suspension of a relatively small portion of the entire batch. The small portion is dried and then ball milled with the required amount of flux to provide a combination activator-flux mixture which can be dry blended with many times its own weight of dry base material. The merit of this method is that large quantities of base material can be fluxed and activated in a short time with very good accuracy.

Although it was well known that activators introduce strong absorption bands which extend into the visible spectrum, this knowledge was not put to practical use until 1952. It has been found that the absorption of light by a ZnS:CdS:Ag phosphor could be reduced by minimizing the activator content. This effect, which produces light-body colored phosphors, is especially notable in P-4 screens where it simultaneously improves light output and decreases screen color nonuniformity.

Firing and Coating

Some inorganic phosphors, such as precipitated calcium tungstate, have a feeble luminescence before they are fired. In general, such luminescence is too weak to be of any practical value, and must be improved by firing at some elevated temperature which may range from 800 to 1250 C. The firing temperature depends on the following factors:

1. Composition of the phosphor
2. The particle size desired
3. The degree of phosphorescence which is desired or tolerable
4. The melting point of the phosphor
5. Such phenomena as solid solution, activation, crystallization, phase change, and solid-state reaction

Phosphors, such as ZnS:Ag or ZnS:CdS:Ag, for black-and-white television can be fired in fused silica boats or crucibles with close-fitting covers to minimize oxidation. This precaution is not sufficient for copper-activated ZnS or ZnS:CdS phosphors which are usually fired at 1200 to 1250 C. These phosphors are fired with blocks of graphite in the small air space between the material and the cover. The graphite prevents excessive oxidation and provides an atmosphere which improves the peak efficiency of the phosphor. This improved result is particulary notable in the case of copper-activated sulfides. Silver-activated ZnS phosphors fired at high temperatures are also improved with a simultaneous narrowing of the emission band. ZnS:CdS:Ag phosphors containing approximately 50 per cent CdS do not seem to be affected at all by the "carbon-firing" technique just described.

The conversion of inert materials or mixtures into phosphors which may be solid solutions or chemical compounds is a solid-state reaction involving time as well as temperature. The controlling factors are the size of the batch, the shape of the charge, and the characteristics of the material being fired. For the sake of homogeneity and to avoid long firing periods, the thickness of the charge should not exceed four inches. The optimum time for a specific result can only be deter-

mined by trial and error.

In addition to the properties controlled by firing, there are several other properties controlled by coating and washing. For ease of handling and blending, a phosphor should be free flowing. It should also have good adherence and resistance to ion burn and photolysis. Unfortunately, at this stage of development, a coating which simultaneously combines maximum adherence with ion-burn resistance is not currently available. For this reason, two coatings have been developed. The first of these is a mixture of hydroxy apatites of the general formula $M_{10}(PO_4)_6(OH)_2$ (where M is calcium, cadmium, or a mixture of both) plus some zinc. This coating provides the phosphor with good wet adherence in the settling process and is satisfactory in tube types where ion burn is not a severe problem. Good burn resistance is obtained by applying what is loosely called a "magnesium silicate" coating. The actual composition of this coating has not been determined, but initially it is probably a mixture of magnesium hydroxide and hydrated silicon dioxide.

The beneficial effects of phosphate coatings were first observed by G. E. Crosby and the author. Later, the phosphate coating was improved, and an ion-burn resistant coating was developed by T. W. Edwards and G. E. Crosby. Their work indicates that the method by which a coating is applied is just as important as the coating material itself.

Soluble chlorides which remain after coating are removed by washing with demineralized water. An arbitrary limit of 0.006 per cent chloride in the finished product has been adopted as standard practice. In the preferred method of washing, a stainless-steel rotary drum filter is used. Washing is followed by air drying at 175 C in a forced-circulation drying oven. Unfired sulfides or any other materials which may evolve hydrogen sulfide should not be dried in the same oven because this gas may react with the coating of the phosphor.

The final step in the manufacture of a sulfide phosphor consists of sifting it through a 250-mesh stainless-steel screen and then storing it in moisture-tight opaque containers until the phosphor has been tested and is ready for use.

OXYGEN-DOMINATED PHOSPHORS

Zinc Orthosilicate: Manganese Activated

For many years zinc orthosilicate (commonly known as willemite) was made without any change of formula. In recent years, its newly acquired importance as the green phosphor for color television has necessitated some changes in its composition and method of preparation.

The formula of one of the materials currently in use is given below:

Zinc oxide	44,000 grams
Silicic anhydride (Special luminescent grade)	20,420 grams
Manganous carbonate	382 grams
Demineralized water	120 liters

The above mixture is milled, dried, crushed, and fired for two hours at 1200 C. After firing, this phosphor is quite soft; it can be easily milled to an average particle size of less than 1 micron and has a peak emission loss of only 10 per cent.

During firing, the solid-state reaction $2\ ZnO + SiO_2 \longrightarrow Zn_2SiO_4$ takes place. As in any chemical reaction, it is desirable to have the reactants as finely divided as possible. Wet milling with 1/2- to 3/4-inch flint pebbles has always been favored as a means of attaining this condition. Manufacturers of ball-milling equipment are experts in establishing the critical factors of mill charge, viscosity, pebble charge, and speed of rotation, and can be very helpful. The milling time is also very important but, for a given set of conditions, the time can be optimized by trial and error. Undermilling does not comminute the ingredients sufficiently to assure efficient compound formation; overmilling tends to contaminate the product with the relatively low-grade silica which abrades from the pebbles. Low peak efficiency caused by incomplete reaction can be corrected by remilling and refiring at 1000 C.

Some applications require a shorter decay than that obtained with the formula given. By means of an eightfold increase in activator concentration, the persistence of Zn_2SiO_4:Mn can be reduced to 25 per cent of its normal value with an attendant decrease of 25 per cent in peak efficiency.

Zinc Orthophosphate: Manganese Activated

Zinc orthophosphate:Mn was first used as the red phosphor in color television tubes by RCA. Originally it was made by the reaction:

$$3\ ZnSO_4 + 2\ NaHPO_4 \longrightarrow Zn_3(PO_4)_2 + 2\ NaSO_4 + H_2SO_4 \qquad (6)$$

A. L. Smith[1] found that zinc orthophosphate:Mn exists in three distinct crystalline forms: the alpha form, which is a light-yellow-green phosphor; the beta or the red form, which peaks at 6380 angstroms; and finally, the orange-red gamma form which peaks at 6320 angstroms. Of these, only the beta form is useful for color television.

S. A. Harper made zinc orthophosphate by the reaction between ammoniacal diammonium phosphate and zinc chloride. He found that the crystal size could be controlled by adjusting the pH of the reaction and he greatly improved the water stability of the phosphor by adding zinc oxide before firing. The original material had a tendency to cake in water suspensions. Caking was due perhaps to the tendency of anhydrous zinc orthophosphate to revert to the tetrahydrate $Zn_3(PO_4)_2 \cdot 4H_2O$. It is not understood how zinc oxide improves the stability of $Zn_3(PO_4)_2$ in water suspensions.

Both of the foregoing methods have been superseded by H. E. McCreary's method which involves the direct reaction between National Formulary (N. F.) grade phosphoric acid and U. S. P. zinc oxide. In McCreary's method, dilute phosphoric acid (specific gravity 1.18) is added at a controlled rate to an aqueous suspension of

zinc oxide containing dissolved manganous chloride. At the end of the reaction the pH should be 2.9 to 3.0. At this point, a stoichiometric excess of 1.93 per cent zinc oxide is added. The reaction produces a mixture of zinc orthophosphate, zinc metaphosphate, and zinc oxide. Firing at 960 C converts the mixture to zinc orthophosphate. The conversion can be attributed to the reaction between zinc metaphosphate and zinc oxide as follows:

$$Zn(PO_3)_2 + 2\ ZnO \rightarrow Zn_3(PO_4)_2 \qquad (7)$$

Apparently an excess of zinc oxide is indispensable for the stability of the phosphor in water suspensions.

An activator concentration of 8.8 mol per cent and a firing of at least 900 C are necessary to form the short-persistence, beta form of zinc orthophosphate:Mn.

REFERENCE

1. Smith, A. L., "Luminescence of the Three Forms of Zinc Orthophosphate:Mn," Journal of the American Electrochemical Society, Vol. 98, No. 9, Sept. 1951

Calculation of Fields and Currents

L. C. Scholz

Harrison

ELECTROSTATIC FIELDS

This chapter discusses the basic relations and laws governing electrostatic fields and currents. The laws are presented axiomatically; for rigor and further clarification, the standard literature[1, 2, 3] should be consulted.

In a static field, the electric field intensity $\vec{E}$ (a vector quantity) may be expressed as the gradient of a scalar potential as follows:

$$\vec{E} = -\text{grad } V = -\Delta V, \tag{1}$$

or

$$\vec{E} = -\left(\vec{i}\,\frac{\delta V}{\delta x} + \vec{j}\,\frac{\delta V}{\delta y} + \vec{k}\,\frac{\delta V}{\delta z}\right) \tag{1a}$$

where $\vec{i}$, $\vec{j}$, and $\vec{k}$ are unit vectors in the x, y, and z directions, respectively. The gradient at a point in space is the maximum space variation of the potential and is defined at that point.

Gauss's law, which expresses a further relation between the electric flux and its source charge, states that the total flux through the surface of a body is equal to the total charge enclosed. This relationship may be written in terms of the flux density $\vec{D}$ (known as the displacement) and the charge density ρ as follows:

$$\iint \vec{D}\, d\,\vec{s} = \iiint \rho\, d\,\sigma, \tag{2}$$

or rephrased in differential form.

$$\frac{\delta Dx}{\delta x} + \frac{\delta Dy}{\delta y} + \frac{\delta D}{\delta z} = \rho \tag{3}$$

From the identity $\vec{D} = \epsilon E$ and a combination of Eqs. (1) and (3), Poisson's equation (in rectangular coordinates) is obtained as follows:

$$\frac{\delta^2 V}{\delta x^2} + \frac{\delta^2 V}{\delta y^2} + \frac{\delta^2 V}{\delta z^2} = -\rho/\epsilon \tag{4}$$

This equation is the basis of the present analysis of electrostatic problems.

A similar result, for cylindrical coordinates, is expressed as follows:

$$\frac{1}{r}\frac{\delta}{\delta r}\left(r\frac{\delta V}{\delta r}\right) + \frac{1}{r^2}\left(\frac{\delta^2 V}{\delta \phi^2}\right) + \frac{\delta^2 V}{\delta z} = \frac{-\rho}{\epsilon} \tag{5}$$

SOLUTIONS FOR SIMPLE GEOMETRIES

In the absence of space charge, Eq. (4) has the form known as LaPlace's equation and may be used for analytic solutions of simple geometries. Although two cases presented below are extremely simple, they provide basic required analysis information.

Infinite Parallel Planes

The potential V at any point z between two parallel planes held at potentials V_1 and V_2, and distance d apart is given as:

$$V = V_1 + \frac{V_2 - V_1}{d}\, z, \tag{6}$$

a linear variation E_z is expressed as:

$$E_z = -\text{grad } V = -\left(\frac{V_2 - V_1}{d}\right) \tag{7}$$

Concentric Cylinders

For concentric cylindrical planes, the basic equation has the following form:

$$V = V_1 + (V_2 - V_1)\,\frac{\ln r/R_1}{\ln R_2/R_1}, \tag{8}$$

and the linear variation is written as

$$E_r = -\frac{V_2 - V_1}{r \ln R_2/R_1} \tag{9}$$

Although analytic solutions are possible for certain other simple geometric configurations, the approximation methods discussed below are generally used.

THE THERMIONIC DIODE

The analysis of the electrical field of a diode differs from the previous problems in electrostatics only in that electrons flow through the interelectrode space. The presence of these electrons results in a charge

(treated as being continuous) which interacts with the total field distribution. In the analysis procedure, Poisson's equation, written with the charge expressed in terms of the current and tube geometry, is integrated over the proper limits.

Before attempting the solution of the equation, however, the more general problem of a current injected into the region between two infinite parallel planes[4, 5, 6] (these planes may be physical electrodes or imaginary in space) must be considered briefly. In the absence of charge, the potential variation, plotted from Eq. (6), is a straight line. As electrons enter the interelectrode space, the potential is everywhere depressed by the field resulting from the charge.

Two possible conditions may exist as the electrons enter the interelectrode space:

(1) an accelerating field may be present as a result of a higher potential on the second plane, as shown in Fig. 1, or, of current.

(2) a retarding field may exist as a result of a higher potential on the first plane, as shown in Fig. 2.

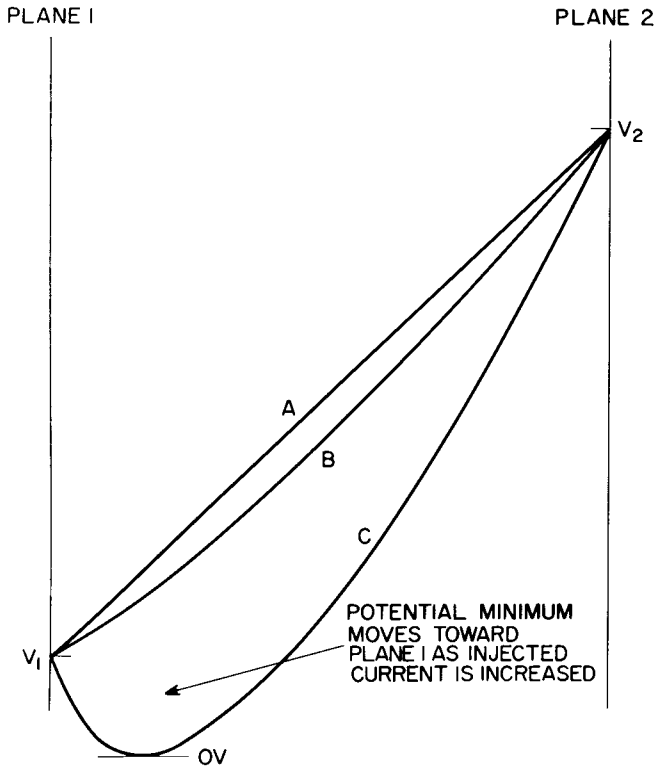


Figure 1. Typical Accelerating Potential Distributions

As the current is increased, a potential minimum is formed; as the current is further increased, the potential at the minimum is lowered to a critical current value at which a transition takes place and a virtual cathode is formed.

In Figs. 1 and 2, curves (A) show the charge-free variation, curves (B) show the effect of a small charge, and curves (C) show the formation of potential minimum.

The concept of self regulation of current, the so-called space-charge-limited current flow, is essential to a clear understanding of the physical processes within an electron tube. Although the idea of a potential minimum is also fundamental and has a physical basis, the sudden transition of potential distribution is less easily visualized and is best explained mathematically. The following analysis of the current flow between two infinite-parallel planes provides a basis for the solution of some of the more important problems of electron-tube design.

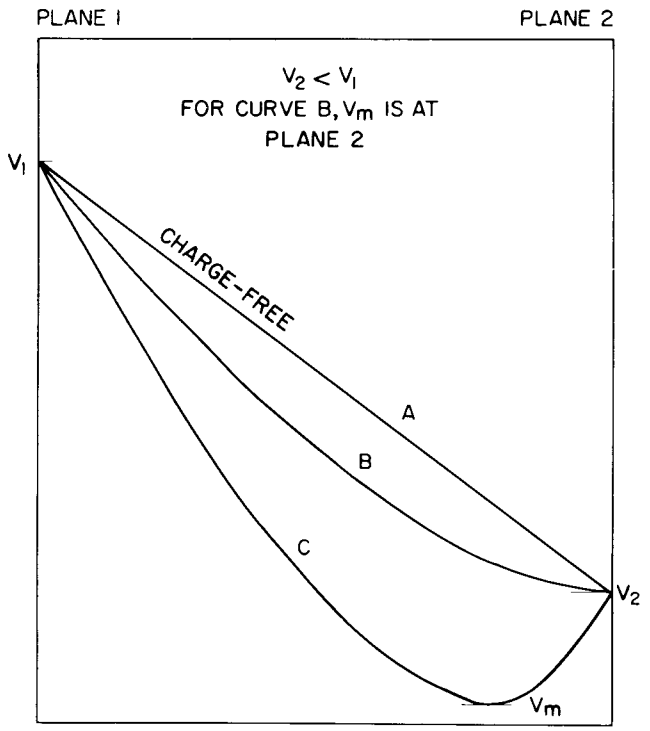


Figure 2. Typical Retarding Potential Distributions

In an infinite-parallel plane geometry,* when a current which has originated at a point of zero potential flows in a Z direction, the velocity in any plane perpendicular to the Z axis is a constant. This condition defines a single-velocity beam. The total current flowing $\vec{J}$ is the sum of the conduction and displacement currents.

$$\vec{J} = \rho\vec{u} + \frac{\epsilon_0 \, \delta\vec{E}}{\delta t} \tag{10}$$

It can be shown that

$$J = \epsilon_0 \frac{dE}{dt} \tag{11}$$

and this expression leads to the following three equations:

*The assumptions inherent in an ideal infinite-parallel plane configuration are outlined in the following section.

$$E = \frac{J}{\epsilon_0} T + E_0 \tag{12}$$

$$U = \frac{e}{m} \frac{J}{\epsilon_0} \frac{T^2}{2} + \frac{e}{m} E_0 T + U_0 \tag{13}$$

$$Z = \frac{e}{m} \frac{J}{\epsilon_0} \frac{T^3}{6} + \frac{e}{m} E_0 \frac{T^2}{2} + U_0 T \tag{14}$$

where E_0 and U_0 are the values at Z = 0, and T is the time of flight of the electrons. These equations may be solved for T by assuming that the value of U_0 is known.

$$T^3\left(\frac{e}{m} \frac{J}{\epsilon_0 12}\right) - \frac{U + U_0}{2} T + Z = 0 \tag{15}$$

Physically realizable solutions for T are positive real values. At most, T has two such roots if the following condition holds true:

$$\frac{1}{12} \frac{e}{m} \frac{J}{\epsilon_0} \leq \frac{1}{54} \frac{(U + U_0)^3}{Z^2}$$

Therefore, the current may have a maximum value

$$J_{max} = \frac{2}{9} \epsilon_0 \frac{m}{e} \frac{(U + U_0)^3}{Z^2} \tag{16}$$

and for this value of current

$$T = \frac{3}{2} \left(\frac{2Z}{U + U_0}\right) \tag{17}$$

The transit time for zero current is given as:

$$T_0 = \frac{2Z}{U + U_0} \tag{18}$$

and

$$T = \frac{3}{2} T_0 \tag{19}$$

Values of current greater than J_{max} result in negative solutions for T because above the value J_{max}, current flow is not stable. Therefore, the equation does not apply above this value. Attempts to increase the current result in changes in the field that turn part of the current back to the source. This effect and related phenomena are observed in pentodes and beam power tubes.

Further discussion of the thermionic diode requires an investigation of the effects of velocity distribution on the equation for space-charge flow. In practical electron tubes, the electrons are initially derived from a heated cathode of metal or some oxide mixture. Thermally emitted electrons travel in a direction perpendicular to the emitting plane at velocities ranging from zero to infinity according to the following Maxwellian distribution function.

$$dN_u = N \frac{m u_1}{KT} \exp\left(-mu_1/2KT\right) du \tag{20}$$

Where dN_u is the number of electrons emitted per second having velocities between u and $u+d_u$, and N is the total number of electrons emitted. This plot of distribution, which has the form shown in Fig. 3, applies to metal emitters as well as the more usual oxide-coated emitter. This spread in velocities contradicts the assumption of a single velocity made above, but for the diode (or grid-cathode region of other tubes), it is possible to compensate for this change by deducing the type of potential distribution that exists in the real diode.

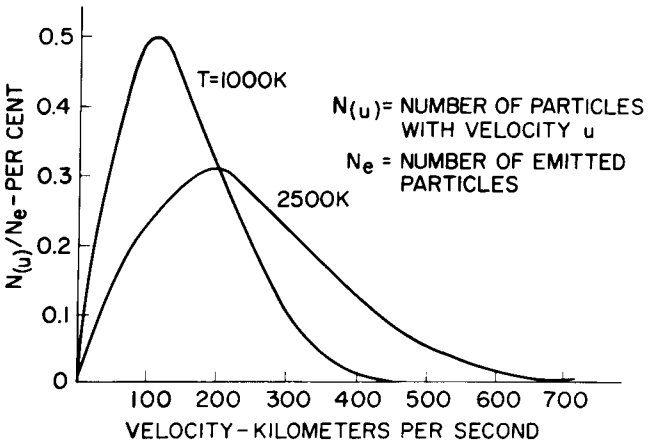


Figure 3. Form of the Normally-Directed Maxwellian Velocity Distribution

In the presence of an accelerating field, a large current can flow in a charge-free tube. The current depresses the potential until a retarding field is formed at the cathode. The retarding field, in turn, limits the current because only electrons having sufficient energy can pass the potential minimum. If enough electrons are available, a limiting condition is established and only a portion of the available electrons reach the anode. This condition is known as space-charge-limited operation. If the current is not sufficient to form a potential minimum, all the available electrons reach the plate, as shown in curve A of Fig. 4. This condition is known as temperature-limited operation. Because all the electrons have energies greater than zero, even zero field at the cathode surface (potential minimum at the cathode surface) permits all the available electrons to flow. These conditions distinguish the boundary between temperature-limited and space-charge limited operation. As the anode potential is reduced, the current is reduced, and the potential minimum moves towards the anode, as shown in curve B of Fig. 4. At some given current, the potential minimum is at the anode surface, as shown in curve C of Fig. 4. A lower current level results in operation in the region of the retarding field. One additional case must be considered: that of electrons emitted with zero energy. If the electrons emitted have zero velocity (energy), the potential minimum must fall on the cathode surface. An accelerating field would cause full current flow and a retarding field would permit no current flow.

THE IDEAL INFINITE PARALLEL-PLANE DIODE

The preceding section presented some of the general

concepts and theory pertaining to the flow of charge between two electrodes. The following is a specific solution for the ideal infinite parallel-plane diode. Because of mathematical complexity, this discussion is restricted to ideal cases. However, experience has shown that these solutions are generally applicable and may be used as a basis for most design problems.

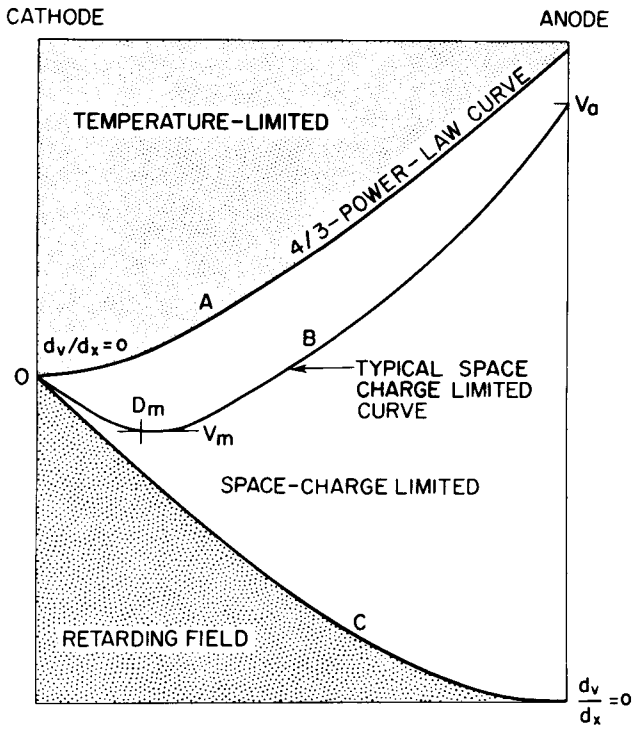


Figure 4. Boundary Potentials Defining Mode of Operation in Diode

The configuration treated here consists of two perfectly parallel uniformly spaced planes infinite in extent. This configuration eliminates the field variation in the directions along the plane and permits a mathematically simple solution. In actual tubes, that are of course finite in extent, end effects must be neglected or some empirical correction must be made. Tubes that have small interelectrode spacing compared to the other dimensions have very small fringe effects. Curved electrode systems may be treated as planes if the spacing is small compared to the curvature.

The method described here is annotated derivation which provides, at each step, the reasons and justification for all assumptions. The discussion is maintained on a general level so that the results may be used as take-off points for less restrictive analyses. It is also very important that the designer realize the limitations and assumptions inherent in the analysis.

Poisson's equation in one rectangular coordinate z is used as the starting point of the analysis.

$$\frac{d^2V}{dz^2} = -\rho(z)/\epsilon_0 \qquad (21)$$

where (z) is a function of z.

Because there are no sources of current in the tube except the cathode, the current must be constant through the tube; the current and space charge density may be related as follows:

$$\rho = J/u \qquad (22)$$

When the principle of conservation of energy is applied to the electrons in an electrostatic field the following relationship between velocity and potential is obtained:

$$U = \left(U_0^2 + \frac{2e}{m} V\right)^{1/2} \qquad (23)$$

where U_0 is the initial z-directed velocity of the electrons as they leave the cathode. Because a distribution function for U_0 leads to a multivalued U(z) that is difficult to manipulate, that case must be treated differently. However, if is is assumed that U_0 is single-valued (zero) and the equation is solved for J, the result is generally useful.

Initial Velocity - Zero

In the most direct analysis procedure, J/u is substituted for ρ in Eq. (21).

$$\frac{d^2V}{dz^2} = \frac{J}{\epsilon_0}\left(\frac{m}{2e}\right)^{1/2} V^{-1/2} \qquad (24)$$

Integration from 0 to z yields

$$E^2 - E_0^2 = \frac{4J}{\epsilon_0}\left(\frac{m}{2e}\right)^{1/2}\left(V^{1/2} - V_0^{1/2}\right) \qquad (25)$$

The boundary conditions must also be defined: E_0 is zero because it has been assumed that U_0 is zero, and the cathode potential V_0 can also be assigned a zero value with no loss in generality. Therefore,

$$E = \frac{dV}{dz} = \left(\frac{4J}{\epsilon_0}\right)^{1/2}\left(\frac{m}{2e}\right)^{1/2} V^{1/2} \qquad (25a)$$

A second integration from 0 to d and 0 to V_a yields

$$\frac{4}{3} V_a^{3/4} = \left(\frac{4J}{\epsilon_0}\right)^{1/2}\left(\frac{m}{2e}\right)^{1/4} d \qquad (26)$$

The significant results, first derived by Langmuir[8] and Childs[9], are as follows:

$$J = 2.336 \times 10^{-6} \frac{V_a^{3/2}}{d^2} \text{ amp per square meter} \qquad (27)$$

$$V = 5.68 \times 10^4 \, J^{2/3} \, z^{4/3} \text{ volts} \qquad (28)$$

$$E = \frac{4}{3} \times 5.68 \times 10^4 \, J^{2/3} \, z^{1/3} \text{ volts per meter} \qquad (29)$$

by simple substitution,

$$E/\text{anode} = \frac{4}{3}\frac{V_a}{d} \qquad (30)$$

This result shows that the electric-field intensity is increased at the anode under space charge conditions because the negative electrons depress the potential (except at the electrodes) and increase the gradient.

The simple relationship shown in Eqs. (27), (28), and (29) is the basis of all tube design. Although a more sophisticated analysis is presented below, much of the work rests on these equations. The three-halves power law is quite general and holds for all geometries (Langmuir[8] and Ivey[10]). The equations may be used to initiate changes in existing tube design or to give a first approximate answer before proceeding with a more refined design.

Initial Velocities - Maxwellian

The distribution of the normally-directed velocities of the emitted electrons is given by the following Maxwellian distribution function:

$$dNu = \frac{Nmu_1}{KT} \exp(-mu_1/2KT)\, du_1$$

A distribution of velocities results in the formation of a potential in front of the cathode, which has a position dm and a magnitude Vm, a negative number. On the anode side of the potential minimum, all electrons move towards the anode and have energies greater than Vm electron volts; on the cathode side, electrons move in both directions. The total emitted electrons (N) move towards the minimum; those with energies less than Vm are returned to the cathode. However, the total current through any plane in the interelectrode space must be a constant. As the electron stream approaches the minimum, the slower electrons are continuously returned, and the faster electrons slowed down. Because the average velocity is constant (Langmuir,[12] p. 423), the following expressions may be used to define the space charge.

On the cathode side of the potential minimum, $0 \le z \le dm$.

$$\rho(z) = -e\left[\int_{(-2\frac{e}{m}Vm)^{1/2}}^{\infty} \frac{dNu}{u} + 2\int_{(2\frac{e}{m}V)^{1/2}}^{(-2\frac{e}{m}Vm)^{1/2}} \frac{dNu}{u}\right] \quad (31)$$

where V is the potential at z, and u the velocity at z is defined by

$$U^2 = U_1^2 + 2eV/m \quad (32)$$

On the anode side of the potential minimum, $dm \le z \le d$

$$\rho(z) = -e\int_{(-2\frac{e}{m}Vm)^{1/2}}^{\infty} \frac{dNu}{U} \quad (33)$$

Under these conditions, the current is determined by the value of the potential minimum and is given by the following expression:[7]

$$J = J_o\, e + \frac{eVm}{KT} \quad (34)$$

where J_o is the total thermionic current available and the exponential e is the electronic charge. This equation is generally used to determine Vm by assuming the values of J_o and J to be known.

A single integration of Poisson's equation gives the following relationship:

$$\left(\frac{dV}{dz}\right)^2 - \left(\frac{dV}{dz}\right)_{dm}^2 = -\frac{2}{\epsilon_o}\int_{Vm}^{V} \rho(2)\, dV \quad (35)$$

where V equals Vm and dV/dz is equal to 0 when z is equal to dm.

The integration of $\rho(z)dV$ results in the combination of some error functions, but a second integration is required to complete the solution. This procedure requires the numerical integration of a new function. Langmuir simplified this evaluation by introducing the following normalizations to give pure numbers.

$$\eta = \frac{e(V - Vm)}{kT} = \frac{11{,}605\,(V - Vm)}{T} \quad (36)$$

$$\xi = \frac{(2m\pi)^{1/4}}{(2KT)^{3/4}}\left(\frac{eJ}{\epsilon_o}\right)^{1/2}(z-dm) = 9.186 \times 10^5 T^{-3/4} J^{1/2}(z-dm) \quad (37)$$

Then

$$\left(\frac{dV}{dz}\right)^2 = \left(\frac{2\xi KT}{e(z-dm)}\right)^2\left[e^{\eta} - 1 \pm e^{\eta}\,\mathrm{erf}(\eta^{1/2}) \mp 2\left(\frac{\eta}{\pi}\right)^{1/2}\right] \quad (38)$$

The upper signs apply in the cathode region, and the lower signs in the anode region;

$$\mathrm{erf}\left(\eta^{1/2}\right) = \frac{2}{\sqrt{\pi}}\int_o^{\eta^{1/2}} e^{-u^2}\, du$$

this expression may be integrated to yield

$$\xi = \int_o^{\eta} \frac{d\eta}{\left[e^{\eta} - 1 \pm e^{\eta}\,\mathrm{erf}\left(\eta^{1/2}\right) \mp 2\left(\frac{\eta}{\pi}\right)^{1/2}\right]} \quad (39)$$

This expression has been evaluated numerically by Kleynen.[14]

Discussion

The preceding analysis is applied to actual design problems through the use of implicit solutions for J. Graphical solutions further facilitate design procedures and aid in the visualization of the effect of the initial velocity on the current.

Under the assumption of zero initial velocity, the current depends on the voltage and spacing only. In the actual case, current is a function of voltage, spacing, cathode temperature, and available cathode current because the position and magnitude of the potential minimum are functions of these variables. Therefore, an adequate analysis must define the effects of these parameters on the current so that a tube may be designed for a given current under a given set of conditions for the greatest utility; this information should be in graphical or tabular form.

The definition of the three important variables η, ξ, and Vm:

$$\eta = \frac{V - Vm}{T} \quad 11{,}605 \tag{40}$$

$$\xi = 9.186 \times 10^5 \tag{41}$$

$$Vm = \frac{kT}{e} \ln \frac{Js}{J} = V_T \ln \frac{Js}{J} \tag{42}$$

By insertion of the proper values for Z and V, the following results are obtained for the cathode and plate:

$$\eta_K = \frac{e}{KT} Vm = \ln \frac{Js}{J} \tag{43}$$

$$Vm = \eta_K V_T = \eta_K \frac{T}{11{,}605} \tag{44}$$

$$\eta_a = \frac{V_a + Vm}{V_T} = \frac{V_a}{V_T} + \eta_K \tag{45}$$

$$V_a = (\eta_a - \eta_K) \frac{T}{11{,}605} \tag{46}$$

$$\xi_K = 9.186 \times 10^5 \; T^{-3/4} \; J^{1/2} \; dm \tag{47}$$

$$dm = \xi_K \; T^{3/4} / 9.186 \times 10^5 \; J^{1/2} \tag{48}$$

$$\xi_a = 9.186 \times 10^5 \; T^{-3/4} \; J^{1/2} \; (d - dm) \tag{49}$$

$$\xi_a = 9.186 \times 10^5 \; T^{-3/4} \; J^{1/2} \; d - \xi_K \tag{49a}$$

Figs. 5, 6, and 7 may be used to determine η_K, ξ_K and ξ_a directly.

These formulas are used, as follows, to design a diode with current density J and voltage drop V:

(1) Assume values for J_S and T_K. (1 to 10 amperes per square centimeter for J_S and 1100 deg K for T_K are practical values for the oxide cathode.)
(2) Calculate (in order) η_K, Vm, and η_a.
(3) Determine from tables or graphs ξ_K and ξ_a.
(4) Calculate d from Eq. (49a).

For example, if J = 40 ma/cm^2 and V = 1.0 volts (J_S = 1000 ma/cm^2).

1. $\eta_K = \ln \frac{1000}{40} = 3.22$

   $Vm = \ln 25 \; (1.1/11.6) = 0.306$ volts

   $\eta_a = \frac{1.0}{0.095} + 3.22 = 13.72$

2. $\xi K_e = 2.26$, $\xi_a = 18$ (from Fig. 6)

3. $d = (\xi_a + \xi_K) \; T^{3/4} / 9.186 \times 10^5 \; J^{1/2} = 0.2$ millimeters or 0.008 inch
   (note that J must be in amperes)

This procedure may be varied to accomodate various combinations of given and desired tube parameters. Some of the actual calculations required in this solution may be eliminated by using the plot of position and magnitude of the potential minimum given in Figs. 8 and 9.

A completely explicit design procedure would provide the current density directly as a function of voltage and spacing; such a plot may be found in Liebermann.[15] A more useful form results through the following normalization.[16] If the current density equals J_1 when the potential minimum falls on the anode surface, J_1 is the transition value as the current density changes from the retarding-field to space-charge-limited mode. If V_1 is that value of anode voltage that gives J_1, J_1 and V_1 are functions of cathode temperature (T_{1C}) and saturation-current density (J_S), but are only slightly dependent on J_S if the ratio J_S/J_1 is very large. Because the ratio is usually large, J_S is assumed infinite and J_1 is designated J_∞. The purpose of this normalization is evident from an examination of a complete potential cross section of a diode, as shown in Fig. 10. The diagram shows that the effective potentials include the work functions of the surfaces. However, because these work functions are not known until after the tube is made, it is necessary to measure or eliminate them. Therefore, after a tube is designed and constructed, J_1 and V_1 are measured and the values are used as a basis for other measurements. If J_1 is not within design limits, something is wrong with the tube. Because the measured values of V_1 include the effect of contact potential, J_∞ is introduced to eliminate contact potential voltages from the measurements. The value of J_1 may be determined from Fig. 11 by letting dm equal d.

An equation for one additional tube parameter, the conductance, is required.

Differentiation of Eq. (27) provides

$$G = \frac{3}{2} \times 2.336 \frac{V^{1/2}}{d^2} \quad \text{micromhos per square centimeter} \tag{50}$$

and

$$\frac{G}{J} = \frac{3}{2} \; V \; \frac{\text{mhos}}{\text{ampere}} \tag{51}$$

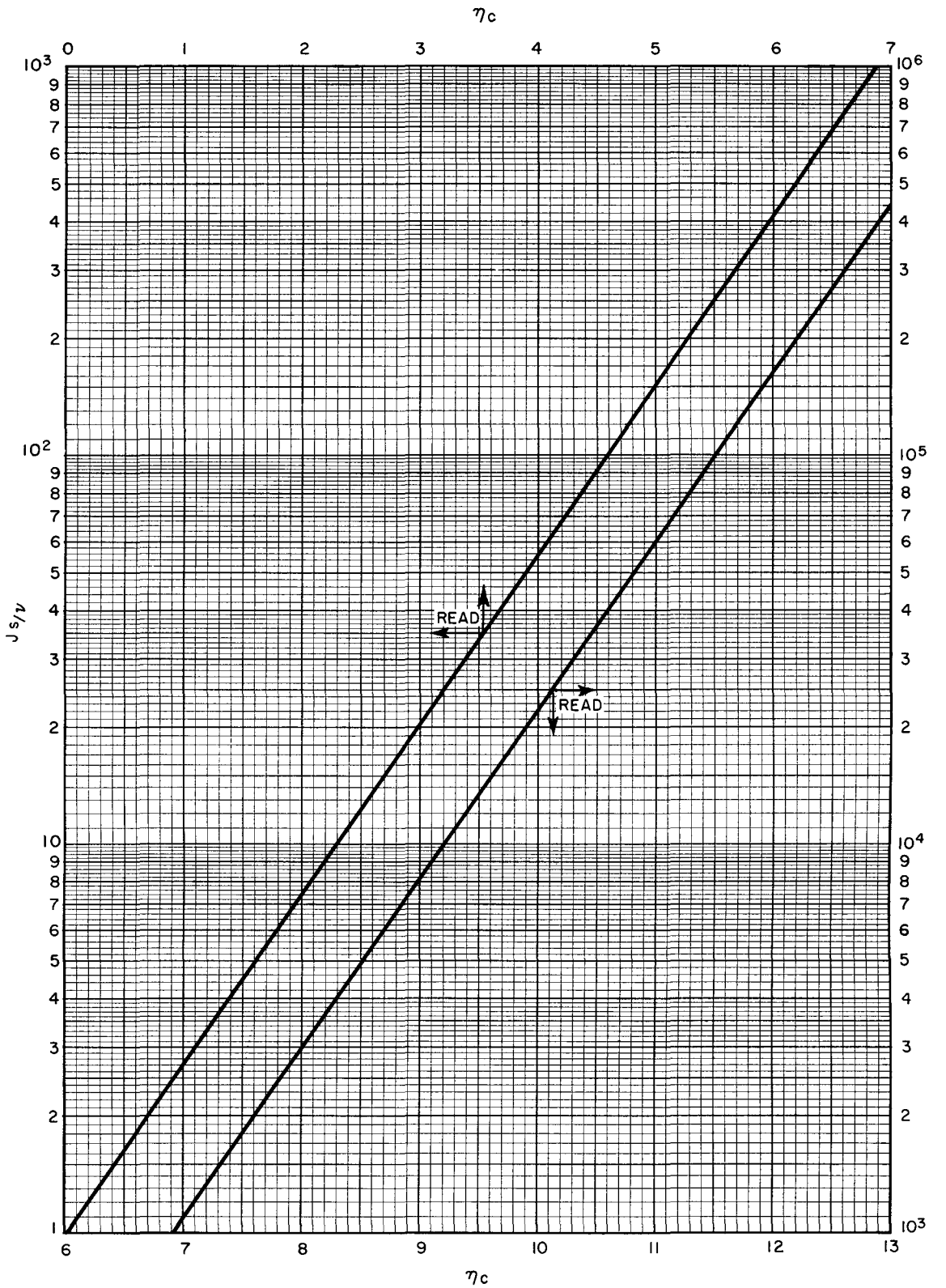


Figure 5. η_c *as a Function of the Ratio* J_s/v

for the case of zero initial velocity. In the exact case, graphical differentiation of the J-V plot is the simplest method, but in the retarding field region

$$J = J_S \exp (eVr/KT) \qquad (52)$$

Therefore,

$$G = \frac{e}{KT} J_S \exp (eVr/KT) \qquad (53)$$

and

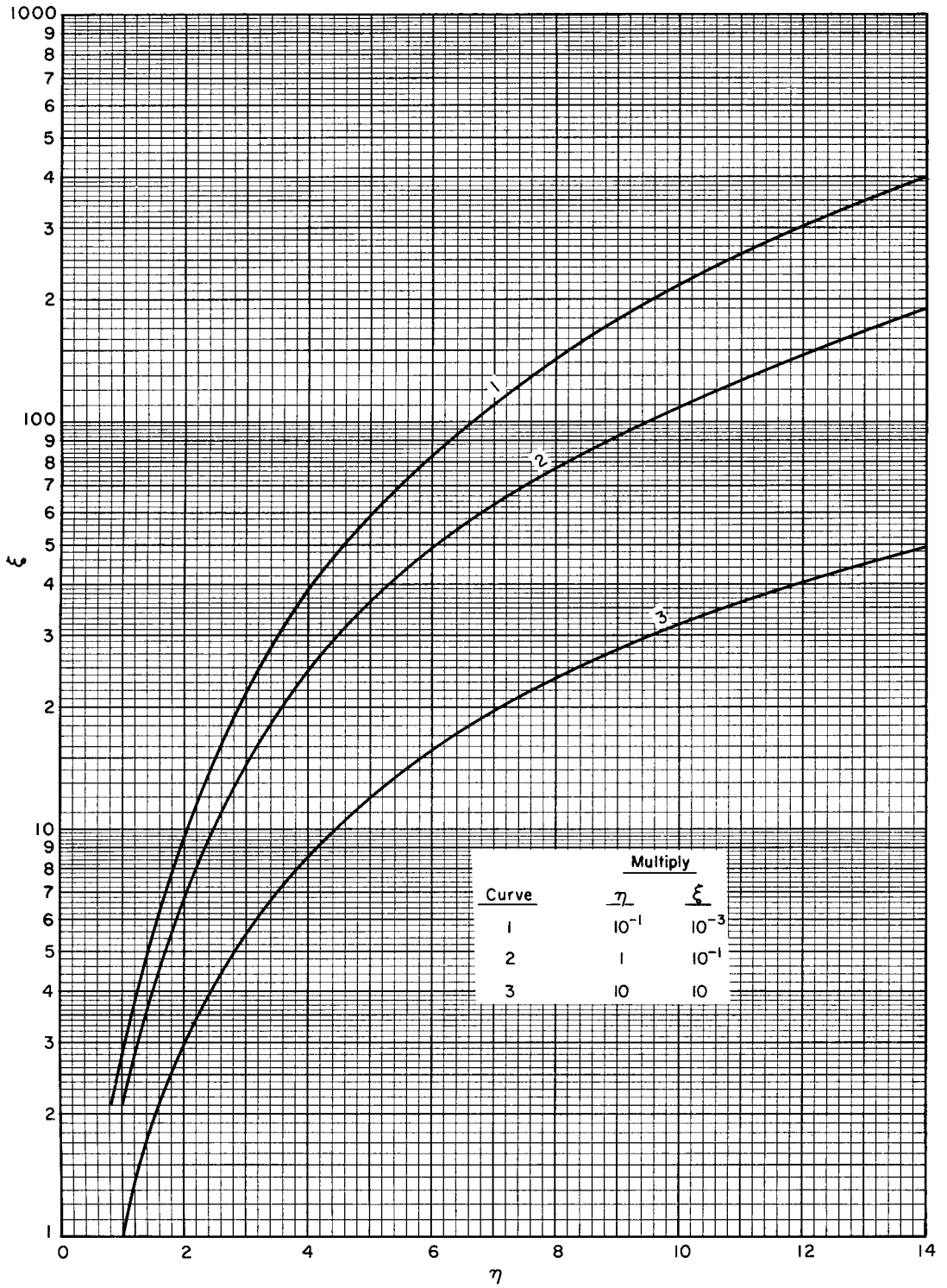


Figure 6. ξ_a *as a Function of* η

$$\frac{G}{J} = \frac{e}{KT} \qquad (53a)$$

Fig. 12 shows the relationship of J, G, and dm at various temperatures and current densities. The set of lines marked (1) is a plot of G as a function of J, when J is less than J_1 (retarding field mode). Set (2) is a plot of dm as a function of J and set (3) is G as a function of J, when J is greater than J_1 and has the specific value of 2 mils.

The figure is used as follows: If a diode has a spacing of X mils, J_1 is found by letting dm equal X. The intersection of line 2 and dm = X yields J_1. Project this to line 1. To the left ($J < J_1$), $G = RJ/KT$ may be found

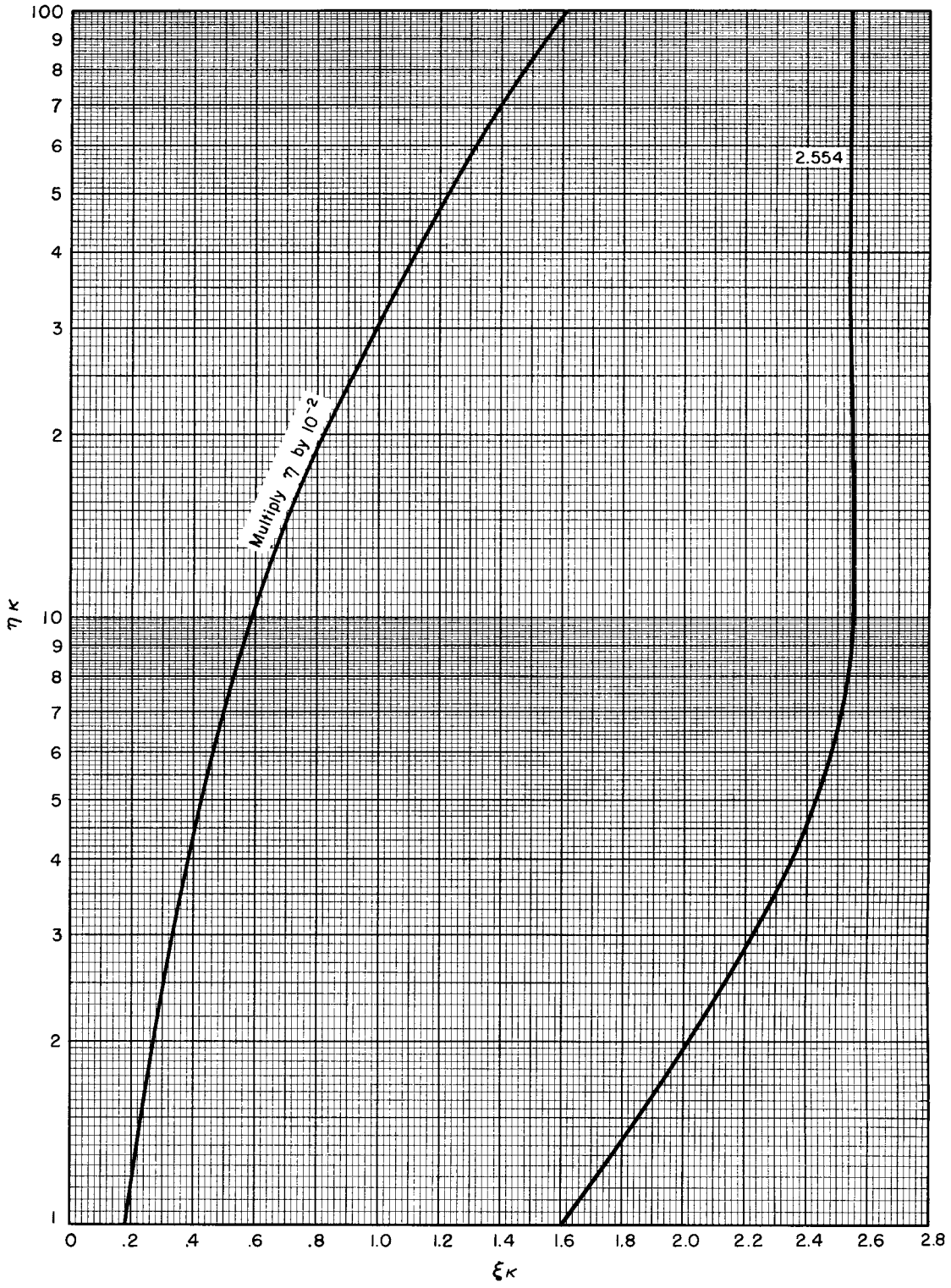


Figure 7. η_K *as a Function of* ξ_K

from line 1; to the right ($J > J_1$), it may be found by graphical differentiation, but the construction of a line parallel to 3 is usually sufficient. The result is a complete plot of G as a function of J; J as a function of V may be calculated. (The designer will probably find it convenient to use universal plots for actual design work.)

The term ξ can be series expanded as shown by Beck,[7] and approximate solutions for J can be obtained.

When only the first term is used,

$$J = \frac{2.336 \times 10^{-6} (V - Vm)^{3/2}}{(d - dm)^2} \text{ a/m}^2 \qquad (54)$$

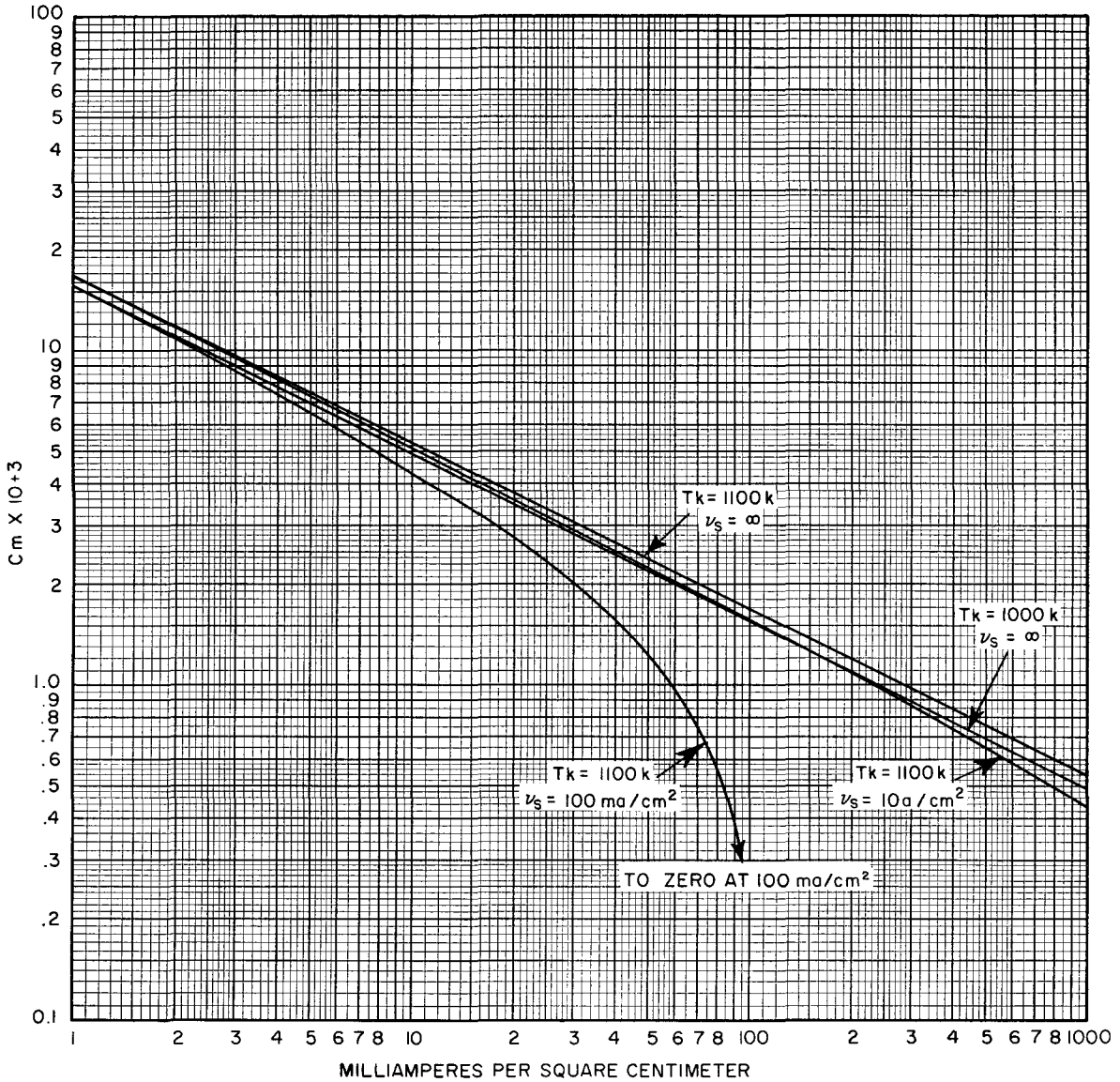


Figure 8. Position of Potential Minimum as a Function of Current Density

When the first two terms are used,

$$J = \frac{2.336 \times 10^{-6} (V - Vm)^{3/2}}{(d - dm)^2} \left[1 + \frac{2.66}{\sqrt{\eta}}\right] \quad (55)$$

The potentials used must include the effects of contact potential.

THE IDEAL CYLINDRICAL DIODE

The following discussion covers only the basic formulation and the results in the analysis of an ideal cylindrical diode because more rigorous examination is not required for the present analysis.

Initial Velocities - Zero

For a given cathode length L,

$$\frac{\delta}{\delta r}\left(r \frac{\delta V}{\delta r}\right) = \frac{I}{2 \pi r \epsilon_o L}\left(\frac{mV}{2e}\right)^{1/2} \quad (56)$$

The solution found by Langmuir[8, 13, 17] is

$$I = \frac{14.66 \times 10^{-6} L V^{3/2}}{r_a \beta^2} \quad (57)$$

$$\text{or } I = \frac{2.336 \times 10^{-6} A V^{3/2}}{r_a r_K \beta^2} \quad (57a)$$

where β^2 is a function of r_a/r_K or r_K/r_a only.

Similar solutions are available for the spherical diode. A spherical diode forms the basis of many convergent electron-gun designs.

Initial Velocity - Maxwellian

The solution to this problem is complicated by the effects of the nonnormal velocities (tangential veloci-

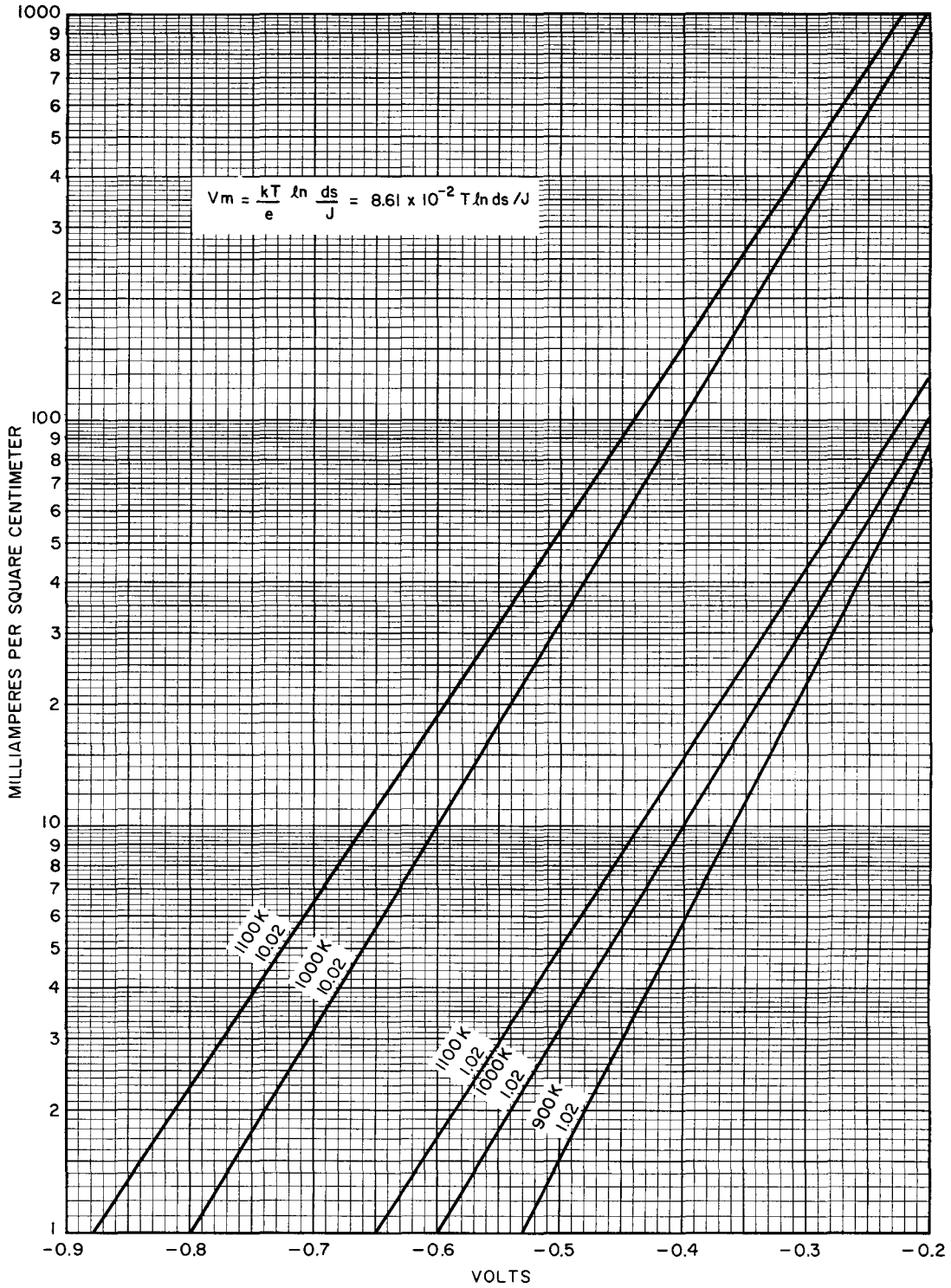


Figure 9. Magnitude of Potential Minimum as a Function of Current Density

ties) on the space-charge distribution, although various approximations have been found.[13, 17] The correction to the three-halves power law is smaller for cylindrical geometry than for the parallel planes. The approximate solution found by including the change in space-charge near the anode is as follows:

$$I = \frac{14.66 \times 10^{-6}\, L}{\beta^2\, ra}\left[V - Vm + \frac{V_o}{4}\left(\ln \frac{V}{\lambda V_o}\right)^2\right]^{3/2} \qquad (58)$$

where

V = the effective voltage

$$V_o = \frac{3}{2}\frac{KT}{e} = \frac{T}{7733} \text{ volts}$$

(The average radial component of initial velocity due to actual radial velocity KT and tangential velocity 1/2 KT.)

Vm = the value of the potential minimum

ra = anode radius

λ = a number between 1 and 2

The position of the minimum does not enter into the equation because it effects only the effective cathode radius which does not enter the equation. Although beta should be determined on the basis of the radius of the potential minimum, such a refinement is not warranted by this approximate solution.

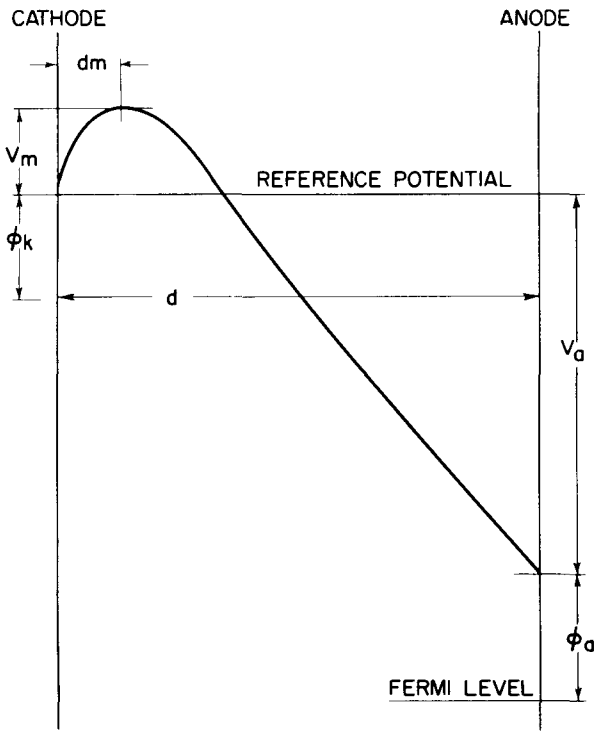


Figure 10. Potential Cross Section of a Thermionic Diode

Discussion

Little need has been found for these formulas, because most close-spaced tubes approximate parallel-plane geometry sufficiently well to justify their use. However, these results can provide the initial steps for the design of electron gun or other such specialized configurations. Fig. 12 shows a graphic solution of Eq. (57) supplied by A. Kauzmann.

THE EXTENSION OF DIODE THEORY TO TRIODES

An analytic solution for the field distribution in a triode is complex, but not impossible if the geometry meets the right criteria. Such a solution is not attempted here because it is not generally useful, but the principle involved in arriving at such a solution will be discussed.

The field distribution around an array of very fine wire was first described by Maxwell[18] (p. 310). If the anode and cathode are not close to the grid, and if the wires are small, this solution may be used for the triode. More involved methods are required to compute the field distribution of a real grid which has wires that are large in diameter compared to their spacing and is close to the cathode and/or anode. In general, these methods involve the replacement of the grid wire by an infinite set of charges, usually a set of multipoles (see Stratton,[3] p. 176). Because of the difficulty in summing the contributions of an infinity of charges, approximations must be used and the solution for the field distribution used indirectly. The quantity wanted is a solution for current flow through the tube with given electrode potentials and geometry. The "equivalent diode" method used is simple, direct, and yields the correct value for the current. It consists of replacing the values of the triode with those of a diode in which anode-cathode voltage and spacing are functions of the geometry and electrode voltages of the triode. The current calculated for the diode is the current for the triode. This method requires that the potential minima (in volts) be the same in each case. Through a unique theorem this may be related to the field intensity (gradient) at various points in the tube.

The Shielding Effect of a Grid

The properties of the field between two electrodes have been discussed in some detail for various geometries, but what are the effects of an array of wires (a grid) placed between the two electrodes? Although a quantitative answer can not be given without reference to a particular geometry, some general results can be discussed by first defining the following parameters:

a - the spacing of the grid wires (for a helical grid the spacing is equal to the pitch)
b - the spacing from the grid to the cathode
c - the spacing from the grid to the plate
r - the radius of the wire

All spacings are measured to the center of the grid wires. The following terms are also used: coverage ratio (2r/a) and window factor (a/b and a/c).

If the cathode and plate are arranged in some arbitrary geometry and have a potential difference of V volts, a flux ϕ exists between them, and field intensities E_1 and E_2 are present at the surface of the cathode and plate, respectively. The potential, flux density, and field intensity at each point in space are known, or can be calculated, and a capacitance coefficient is assigned to the two electrodes. If a third electrode (say a grid) is placed in the field, it intercepts some of the flux and alters the field distribution. If it is initially charged, its field interacts with the field of the other electrodes, and the gradient, flux density, and potential at each point in space are no longer the same. The magnitude of the change depends on the shape of the third electrode and how and where it is placed in the field. An interposed electrode that physically shields one element from the other has a greater effect than one placed near the edge of the field. The quantity mu defines the shielding effect of a grid in an electron tube.

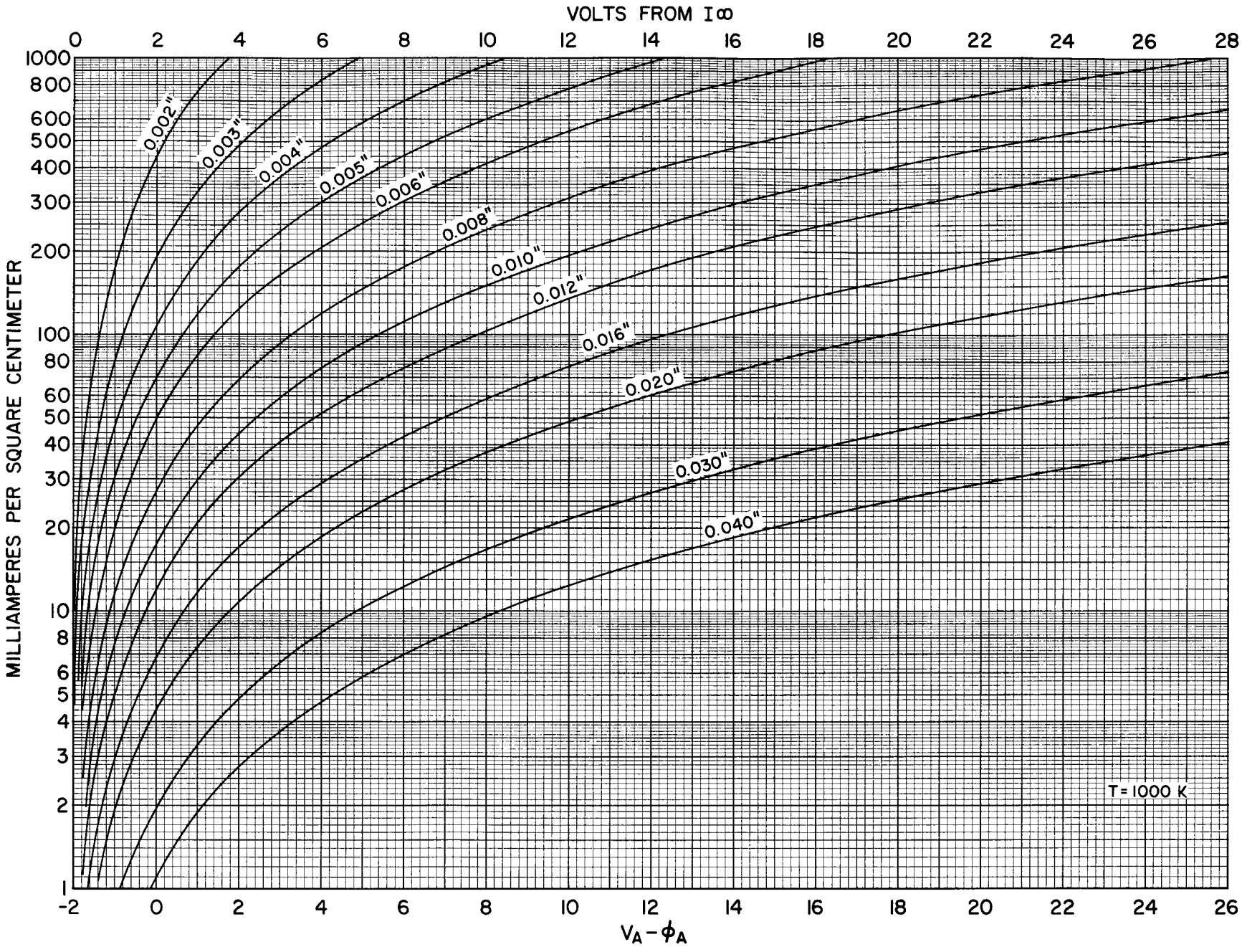


Figure 11. Plot of J vs. V with d as a Parameter

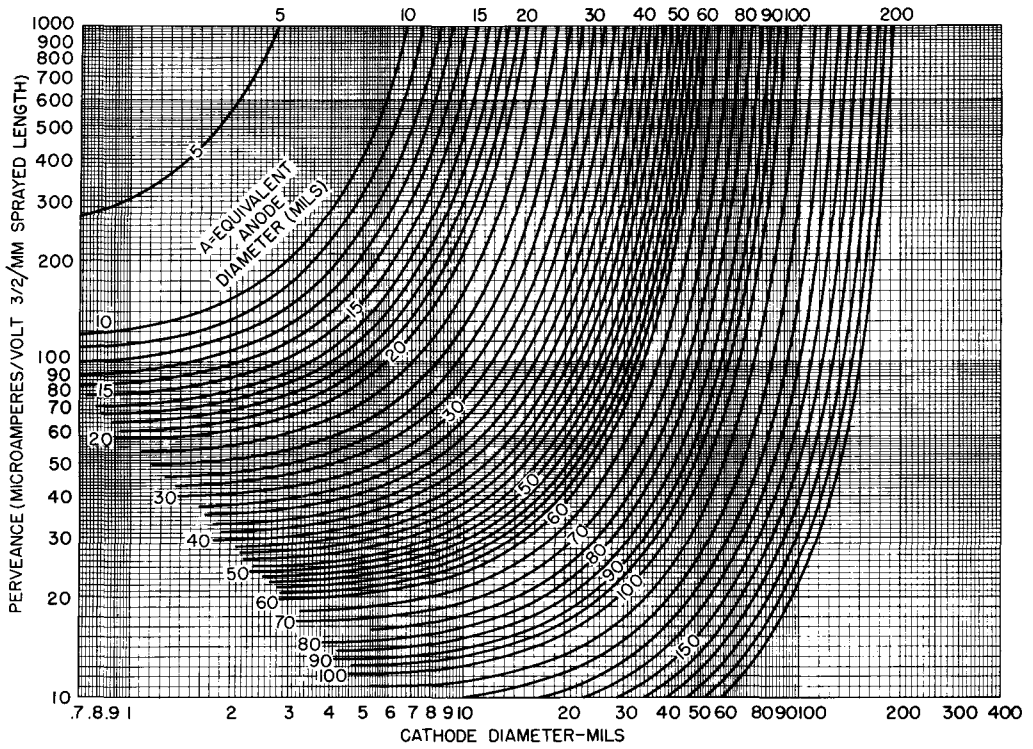


Figure 12. Universal Plane-Parallel Diode Design Curves

The following definitions of mu represent two separate points of view, but give the same result.

From the user's viewpoint, the tube is considered as a device (a three-terminal network), of which the internal workings are not known, and mu is defined as "the limit of the ratio of the change in anode voltage to the change in grid voltage with the anode current held constant." This relationship is expressed as follows:

$$\mu = \frac{\delta Va}{\delta Vg}\Bigg|_{Ia = \text{constant}} \tag{59}$$

From the designer's point of view, the tube is considered as a problem in electrostatics. The fields due to the three electrodes are considered, and mu is defined as "the limit of the change in the cathode-surface charge density due to a change in grid voltage to the change in cathode-surface charge density due to a change in anode voltage." This relationship may be expressed mathematically as follows:

$$\underset{\Delta \to 0}{\text{LIMIT}} \quad \frac{\Delta \sigma K / \Delta Vg}{\Delta \sigma K / \Delta Va} = \mu \tag{60}$$

At the boundary of a conductor

$$\sigma_K = E \text{ normal}$$

This term may be substituted in Eq. (60) as follows:

$$\underset{\Delta \to 0}{\text{LIMIT}} \quad \frac{\Delta E_{NK} / \Delta Vg}{\Delta E_{NK} / \Delta Va} = \mu \tag{60a}$$

and reduced to

$$\mu = \frac{\sigma Va}{\sigma Vg}\Bigg|_{\delta K = E_{NK} = \text{constant}} \tag{60b}$$

Although this is an electrostatic definition, it can also be applied when space charge is present. The equivalency of the two definitions may be shown from an examination of the field in a diode having a flow of electrons which have initial velocities; a discrete value for the gradient is present at the cathode surface for each value of current. Hence, the current is maintained constant by maintaining the off-cathode field constant. It should be understood that in defining mu (electrostatically) it is not necessary to refer to the cathode surface; any point in space may be used. The cathode was used in this case to illustrate the equivalency of the two definitions. In addition, it is not necessary to take the limit in the electrostatic case, because the mathematics is linear and finite deviations may be used. In the space-charge-considered case, because the current is not a linear function of voltage, the limiting value must be used.

The two charts shown in Figs. 13 and 14 may be used to calculate the penetration factor (Durchgriff) of any grid in a plane parallel geometry. Penetration factor is the inverse of mu and is sometimes given as a percentage. These charts are from an article by Kleijnen.[19] An additional chart with somewhat more limited range is shown in Figs. 15 and 16. This chart is useful in calculating mu for tubes having cylindrical geometry. Various methods are discussed below for measuring mu of odd structures by analog methods. The results of a set of such measurements made by Hsu and Horton[20]

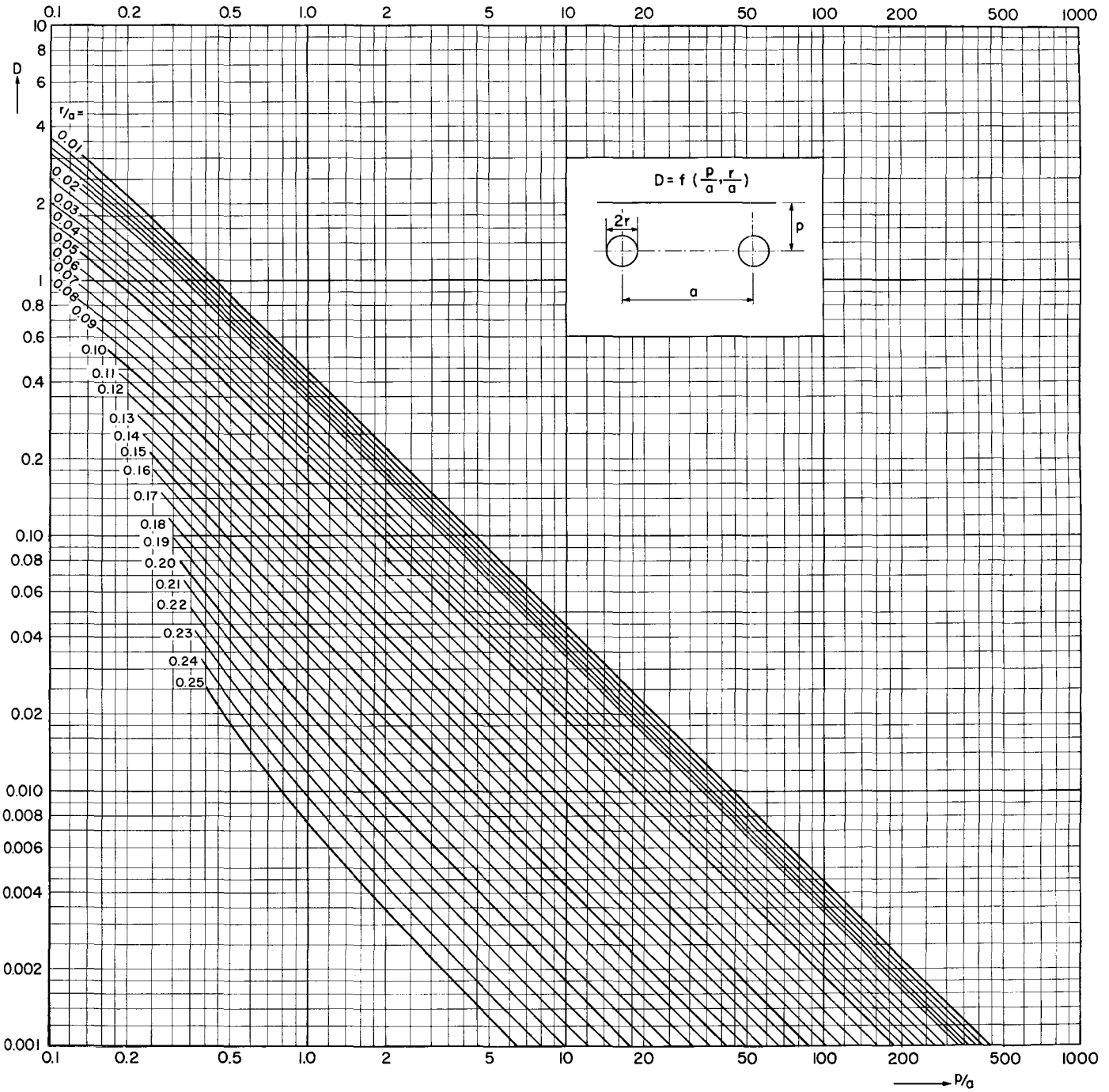


Figure 13. Design Data for Cylindrical Diodes

are shown in Fig. 17. These measurements may be used to determine the mu of various mesh grids.

Various calculations have been made of mu; Maxwell[18] probably made the first such calculation. His results may be used for grids which have small wires and are placed far from any other electrode. The method replaces the grid with an array of infinitesimal line charges which have potential lines that are circles near the charge. The charges are chosen such that the resultant potential lines fit the grid wires. Farther from the charges, the potential contours become cardioidal and can no longer be fitted to a round grid wire. Vodges and Elder[21] overcame this difficulty by replacing the simple line charge with a dipole. This method resulted in a better fit for wire diameters up to 16 per cent of the grid to cathode spacing. Herne[22] replaced the round grid wire with a wire having a polygon cross section and used a Schwartz-Christofel transformation to obtain results that are valid for cases in which the anode is close to the grid plane. Ollendorf[23] proceeded further, and he obtained a multiplicity of charges by taking successive approximations that included more and more terms. His third and fourth approximations are useful for large grid wires up to 0.72 2r/a, but not for closely-spaced anodes. Kleijnen replaced each grid

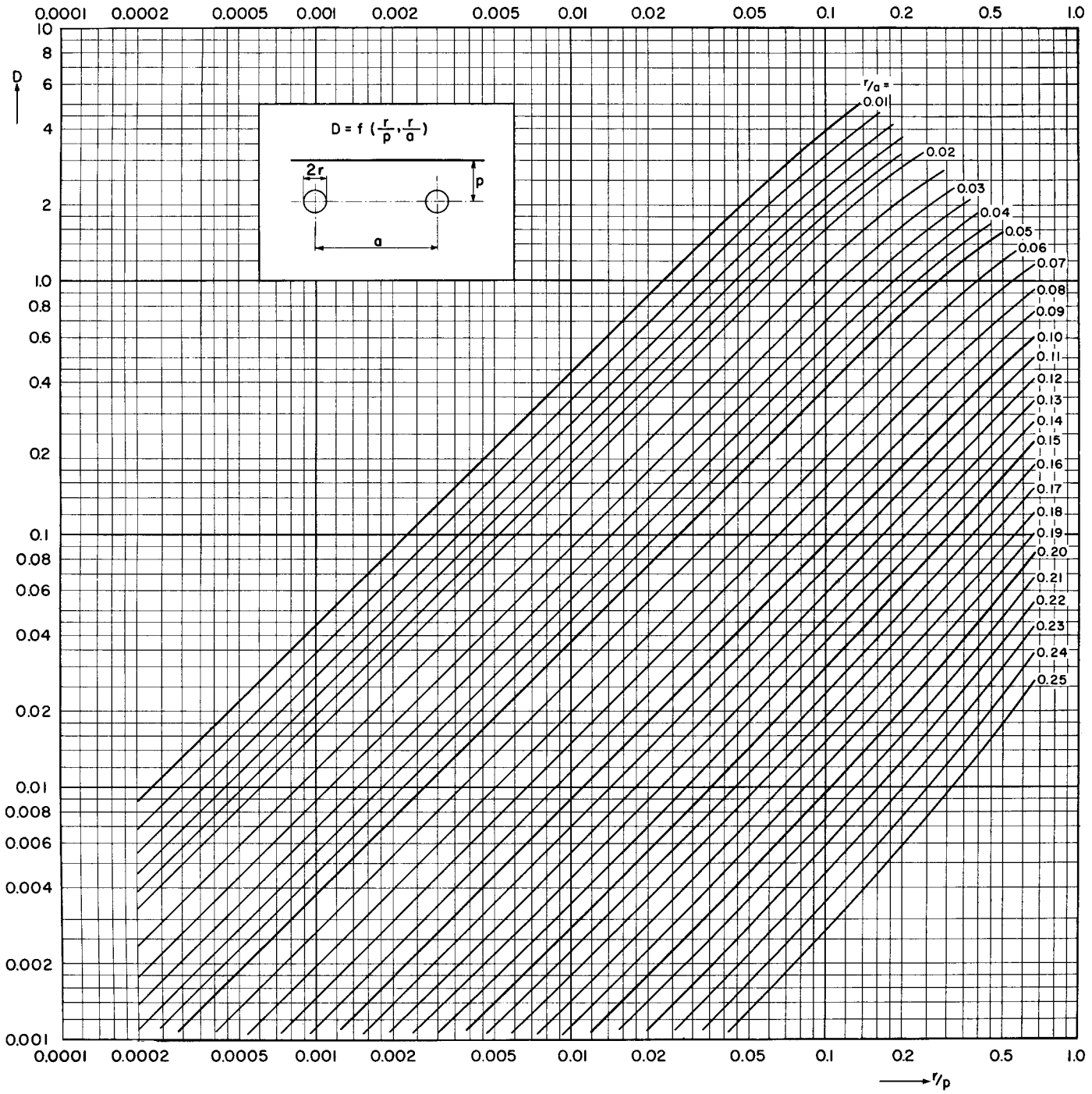


Figure 14. Penetration Factor (Kleijnen)

wire by a series of n-pole charges where n varies from one to infinity. His results, presented below, are theoretically good for all coverage ratios and for all grid-anode spacings up to the limiting values of 2r = a and s = a; s is the grid-anode spacing. The results discussed above are not valid for tubes in which the pitch is large compared to the grid-cathode spacing. This mathematically special case (which includes most modern high-performance tubes) is discussed below. Under these conditions, the field at the cathode surface is not smooth (i. e., it varies from a minimum at a point under a grid wire to a maximum at a point between grid wires) and, as a result, the current density is not uniformly distributed. The present discussion is restricted to the case in which the grid-cathode spacing is larger than the pitch.

Calculation of Equivalent Potentials

The importance of a unified approach to the electronics of a vacuum tube is most important at high frequencies when the tube capacitance and inductance are a significant part of the circuit. The interelectrode capacitances of a tube, however, are not constant, but

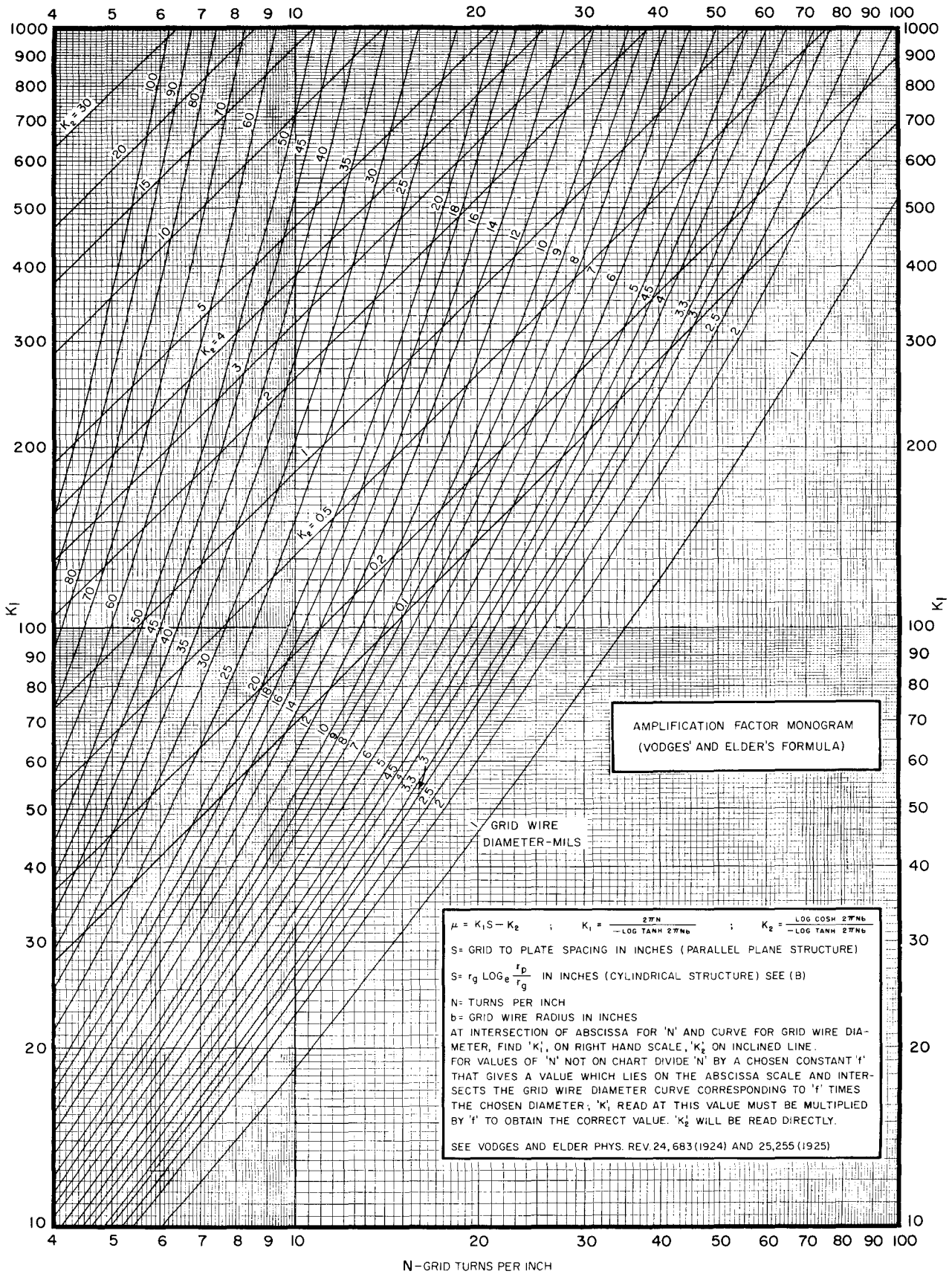

AMPLIFICATION FACTOR MONOGRAM
(VODGES' AND ELDER'S FORMULA)
GRID WIRE
DIAMETER-MILS
$\mu = K_1 S - K_2$; $K_1 = \frac{2\pi N}{-\text{LOG TANH } 2\pi Nb}$; $K_2 = \frac{\text{LOG COSH } 2\pi Nb}{-\text{LOG TANH } 2\pi Nb}$
S= GRID TO PLATE SPACING IN INCHES (PARALLEL PLANE STRUCTURE)
$S = r_g \text{ LOG}_e \frac{r_p}{r_g}$ IN INCHES (CYLINDRICAL STRUCTURE) SEE (B)
N= TURNS PER INCH
b= GRID WIRE RADIUS IN INCHES
AT INTERSECTION OF ABSCISSA FOR 'N' AND CURVE FOR GRID WIRE DIAMETER, FIND 'K_1', ON RIGHT HAND SCALE, 'K_2' ON INCLINED LINE.
FOR VALUES OF 'N' NOT ON CHART DIVIDE 'N' BY A CHOSEN CONSTANT 'f' THAT GIVES A VALUE WHICH LIES ON THE ABSCISSA SCALE AND INTERSECTS THE GRID WIRE DIAMETER CURVE CORRESPONDING TO 'f' TIMES THE CHOSEN DIAMETER; 'K_1' READ AT THIS VALUE MUST BE MULTIPLIED BY 'f' TO OBTAIN THE CORRECT VALUE. 'K_2' WILL BE READ DIRECTLY.
SEE VODGES AND ELDER PHYS. REV. 24, 683 (1924) AND 25,255 (1925)
K_1
N–GRID TURNS PER INCH

Figure 15. Penetration Factor (Kleijnen)

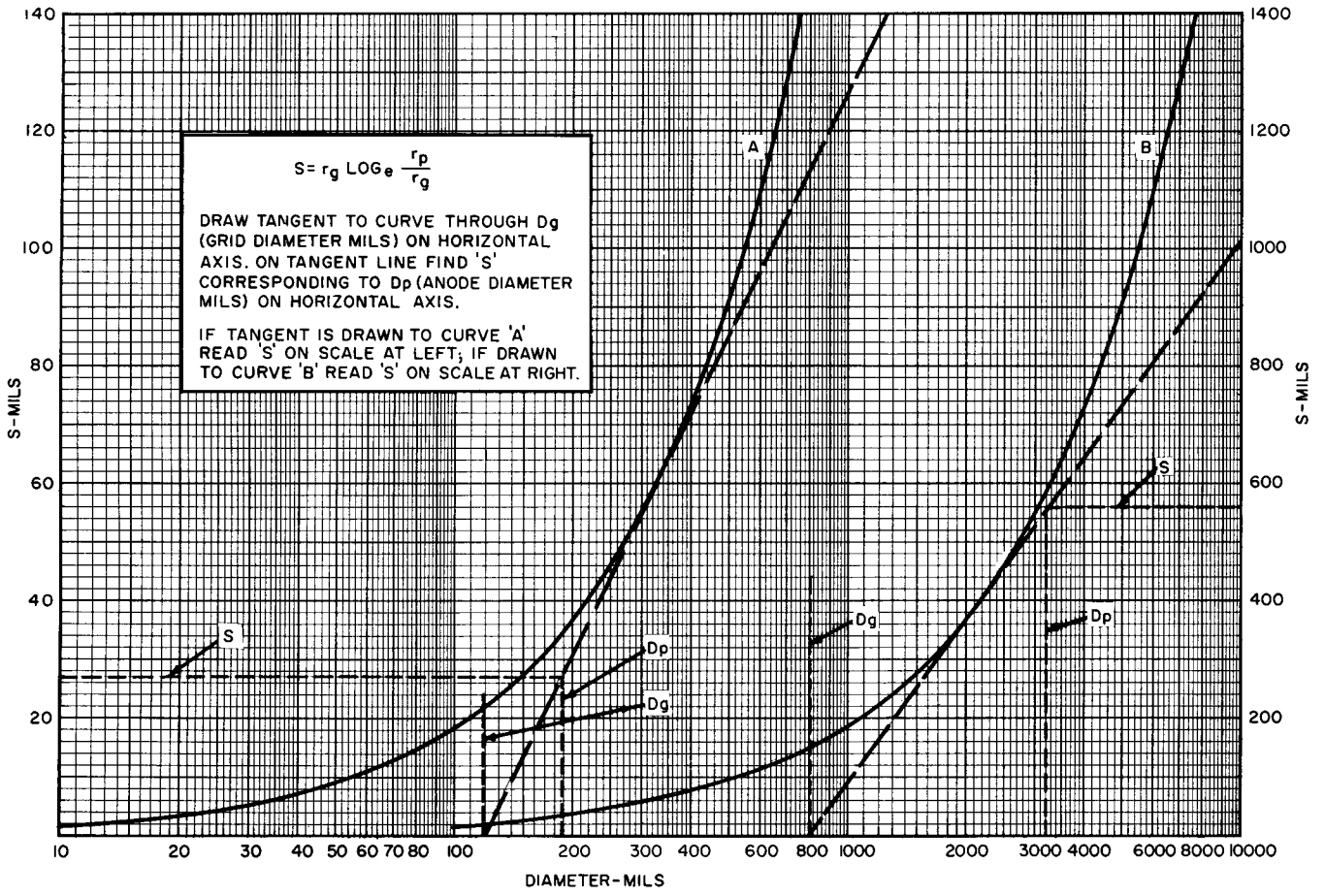


Figure 16. Amplification Factor Nomogram (Vodges & Elder Formula)

vary with the amount of space charge present. At high frequencies, electrons take an appreciable part of a cycle to cross the grid-cathode space and hence do not encounter a static conservative field. At these high frequencies, the electrons absorb energy, manifested as a power requirement for the grid drive. An effective design procedure for these tubes must treat the tube and circuit together as has been done in more recent years by Llewellyn and Peterson.[27] The earlier approach is more simple, but less accurate.

The most simple approach to the calculation of equivalent diodes[24, 25, 26] is purely electrostatic and only charge-free regions are considered. The first refinement, a correction for space charge in the grid-cathode region, is commonly used. Space charge in the grid anode region is very rarely considered because it leads to undue difficulty. However, because even small amounts of space charge may have large effects, an analysis has been made, as a second order correction, that includes some effects of space charge in the anode region, but not the space charge itself. This approach, although similar to the Llewellyn-Peterson approach, includes only the dc terms.

The Simple Charge-Free Model is a basic analysis approach which leads to the standard form for the equivalent-diode voltage. Although the result is widely used, it is not accurate for high-frequency or low-voltage tubes (i. e., those having a large transit time).

The tube is replaced by the electrostatic capacitances between elements (these capacitances cannot be measured because they include only those active portions of the electrodes that actually carry current), see Fig. 18. This delta or pi network is transformed to a tee, by the ordinary rules of circuit analysis. The transformation equations are as follows:

$$C_A = P/C_{GK} \tag{61a}$$

$$C_G = P/C_{AK} \tag{61b}$$

$$C_K = P/C_{GA} \tag{61c}$$

where $P = C_{GK}\,C_{AK} + C_{GA}\,C_{AK} + C_{GA}\,C_{GK}$

It is assumed that there is an imaginary plane in the tube at the mode (n) for which potential v is determined.

The equation VC = Q relates the charge on the electrodes to the potential difference between them. In addition, because the plane at n is in space and no real

net charge may exist there, the charge at n has the following identity:

$$0 = (V_g - V)\,C_g + (V_a - V)\,C_a + (V_K - V)\,C_K,$$

V may be found to be

$$V = \frac{V_g\,C_g + V_a\,C_a + V_K\,C_K}{C_g + C_a + C_K} \tag{62}$$

$$V = \frac{V_g + V_a\,\frac{C_a}{C_g} + V_K\,\frac{C_K}{C_g}}{1 + \frac{C_a}{C_g} + \frac{C_K}{C_g}} \tag{63}$$

From Eq. (61),

$$C_a/C_g = C_{aK}/C_{gK}, \text{ and } C_K/C_g = C_{aK}/C_{ga}$$

This relationship indicates that the imaginary plane is situated in the plane of the actual grid. From our definition of mu, Eq. (60), the relationship between the capacitances is expressed as follows:

$$C_g/C_a = \mu_{ga}, \tag{64}$$

and

$$C_g/C_K = \mu_{gK} \tag{65}$$

Eq. (63) may be rewritten as follows:

$$V = \frac{V_g + \frac{V_g}{\mu(g\text{-}a)} + \frac{V_K}{\mu(g\text{-}K)}}{1 + \frac{1}{\mu(g\text{-}a)} + \frac{1}{\mu(g\text{-}K)}} \tag{66a}$$

The mu of a helical grid is a function of wire size, pitch, and spacing to the plane under consideration. For a grid between two planes, as in a triode (with $V_K = 0$), the relationship is expressed as follows:

$$V = \frac{V_g + \frac{V_a}{\mu}}{1 + \frac{1}{\mu}\left(1 + \frac{s}{b}\right)} \tag{66b}$$

where $\mu = \mu_{ag}$, b = g-K spacing, and s = g-a spacing.

The preceding discussion has been limited to the charge-free case. The following correction considers the charge in the cathode-grid region (called space 1), but neglects the grid-anode region (space 2). If it is again assumed that the simplified case is based on a cathode that emits an infinity of electrons at zero velocity, we may express the gradient on the cathode side of the grid plane as follows:

$$E_{g1} = \frac{4}{3}\frac{V}{b} \tag{67}$$

This value is 4/3 of the gradient in the charge-free model. The same gradient is achieved in the model and the actual tube by reducing the spacing in the model to 3/4b. This reduces the gradients the same but also reduces the reverse mu to 3/4 of its calculated value. Therefore, Eq. (66) may be rewritten as follows:

$$V = \frac{V_g + \frac{V_a}{\mu}}{1 + \frac{1}{\mu}\left(1 + \frac{4}{3}\frac{s}{b}\right)} \tag{68}$$

This equation represents the charge-free model with a first-order correction for charge in the cathode space.

The simplified two-space analysis of a triode attempts to include the effects of space charge in both spaces. The final result includes the effect of the finite transit time, but neglects the space charge itself. First, an expression for the field in a gap with current flowing is obtained which is then applied to the various regions in the triode. The essential steps of the procedure are presented below. Space 1 is the grid-cathode region and space 2 is the grid-anode region. A zero subscript refers to charge-free values. The analysis applies to a single-velocity current flow.

The electric-field intensity at the entrance plane is expressed as follows:

$$E_1 = \frac{V_2 - V_1}{Z_2 - Z_1} \cdot \frac{T_o}{T} - \frac{J}{2\epsilon_o}\,T \tag{69}$$

and at the exit plane is expressed as:

$$E_2 = \frac{V_2 - V_1}{Z_2 - Z_1} \cdot \frac{T_o}{T} + \frac{J}{2\epsilon_o}\,T \tag{70}$$

These values are the linear gradients modified by a space-charge factor.

This general result may be applied to a triode by evaluating the gradients at each electrode in each space. Space 1 is the g-K gap and space 2 is the g-a gap. The grid is a common element and has a gradient on each side. The numerical subscripts refer to the space and the zero indicates no space charge.

In the grid-cathode space:

$$E_K = 0 \tag{71}$$

$$E_{g1} = \frac{4}{3}\,\frac{V}{b} \tag{72}$$

$$T_1 = \frac{4}{3}\,\frac{\epsilon_o}{J}\,\frac{V}{b} \tag{73}$$

These expressions have been previously derived.

In the grid-anode space:

$$E_a = \frac{V_a - V}{s}\,\frac{T_{20}}{T_2} + \frac{2}{3}\,\frac{V}{b}\,\frac{T_{20}}{T_2} \cdot \frac{T_2}{T_{20}} \tag{74}$$

$$E_{g2} = \frac{V_a - V}{s}\,\frac{T_{20}}{T_2} - \frac{2}{3}\,\frac{V}{b}\,\frac{T_{20}}{T_2} \cdot \frac{T_2}{T_{20}} \tag{75}$$

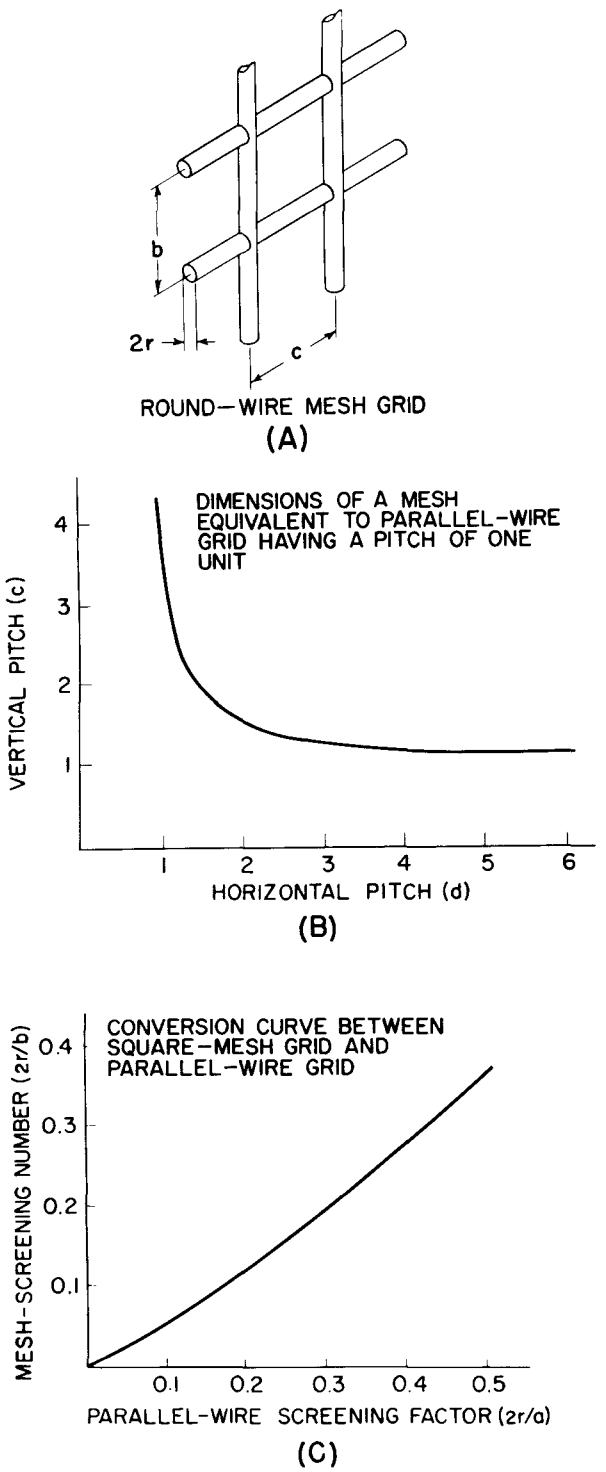


Figure 17. Relations between Helical and Mesh Grids: (A) Round-wire mesh grid; (B) Horizontal pitch (d); (C) Parallel-wire screening factor (2r/a)

These expressions can be used by evaluating the transit time ratios as follows:

$$\frac{T_{20}}{T_1} = \frac{2}{3}\,\frac{s}{b}\,\frac{V^{1/2}}{V_a^{1/2} + V^{1/2}} \qquad (76)$$

$$\frac{T_2}{T_{20}} = \left(1 + \frac{4}{27}\,\frac{J}{J_s}\right) \qquad (77)$$

An expression for the equivalent voltage V is obtained from these equations as follows:

$$V = \frac{V_g + V_a / \mu'}{1 + \frac{1}{\mu'}\left[1 + \frac{4}{3}\,\frac{s}{b}\,\frac{T_2}{T_{20}}\left(1 + \frac{1}{2}\,\frac{T_2}{T_1}\right)\right]} \qquad (78)$$

where

$$\mu' = \mu\,\frac{T_2}{T_{20}}$$

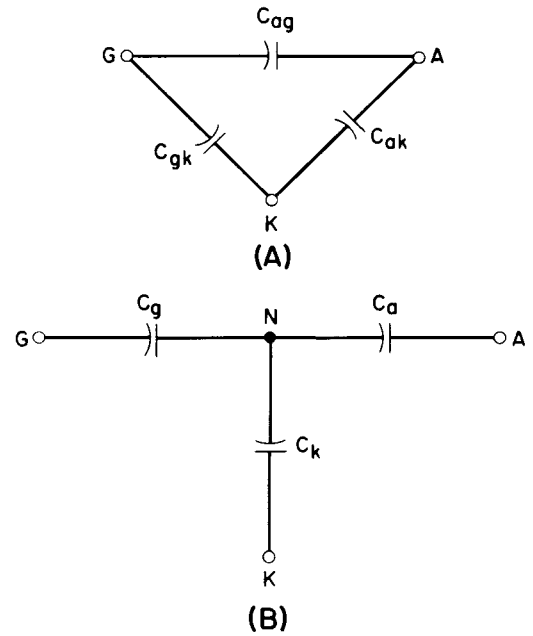


Figure 18. Electrostatic (Capacitance) Representation of a Triode: (A) Delta representation of interelectrode capacitances; (B) Tee representation of interelectrode capacitances (transformed from Delta)

If T_2 is zero and T_2/T_{20} approaches 1, the expression can be reduced to the previous result [Eq. (68)]. However, if T_2/T_{20} approaches 1 when T_2 is not zero, the result is a solution which may be said to be a first-order correction for space charge in the anode space. It simply takes into account the finite transit times of the electrons and adds a small correction to the field due to the accumulated charge.

As a final result,

$$V = \frac{V_g + V_a / \mu}{1 + \frac{1}{\mu}\left[1 + \frac{4}{3}\,\frac{s}{b}\left(1 + \frac{1}{2}\,\frac{T_{20}}{T_1}\right)\right]} \qquad (79)$$

and

$$G_m = \frac{G}{1 + \frac{1}{\mu}\left[1 + \frac{4}{3}\frac{s}{b}\left(\frac{2}{3} + \frac{1}{2}\,\frac{T_{20}}{T_1}\,\frac{V^{1/2} + 3/2 V_a^{1/2}}{V^{1/2} + V_a^{1/2}}\right)\right]} \qquad (80)$$

The ratio T_{20}/T_1 may be evaluated from Eq. (81) by assuming some small value for V consistent with the known requirements of the design.

Interpretation

This quasi-static representation for an ideal triode, which includes at least first order corrections for the space charge in both the cathode and anode spaces, can be used for design or analysis. For accurate calculations, the contact potentials, potential minimum, and the electrode potentials must be included. Because the determination of the contact potential is difficult and accurate prediction of current is almost impossible, the design of electron tubes is somewhat empirical. Any methods of analysis should consider information on the accuracy of design, an estimate of degradation in performance from imperfect grids and parts, and information about the contact potential. This information may be used to determine: (1) errors in parts and spacing, (2) the actual electrode voltages required to obtain a desired current level, and (3) any improvements which may be effected by changes in grid design.

An analysis procedure for triodes (and multigrid tubes) is derived from that previously discussed for the diode. The following tests* may be performed on a General Radio vacuum-tube parameter bridge.

Normally, for any finite grid-wire size, the cathode current is bunched toward the center of the grid aperture (see Fig. 19); a diode of similar dimensions would have laminar flow. Diode conditions may be duplicated by adjusting the electrode voltages until laminar flow results. Under these conditions, the grid intercepts 2r/a per cent of the cathode current and the space around it must be at the same potential;** no change would result if the grid were removed. The grid is said to be at natural or space-charge potential.

In the actual case, shown in Fig. 19D, the potential line that is at grid potential just grazes the lower edge of the grid wires. The diode spacing in this case is (b-r) rather than b.

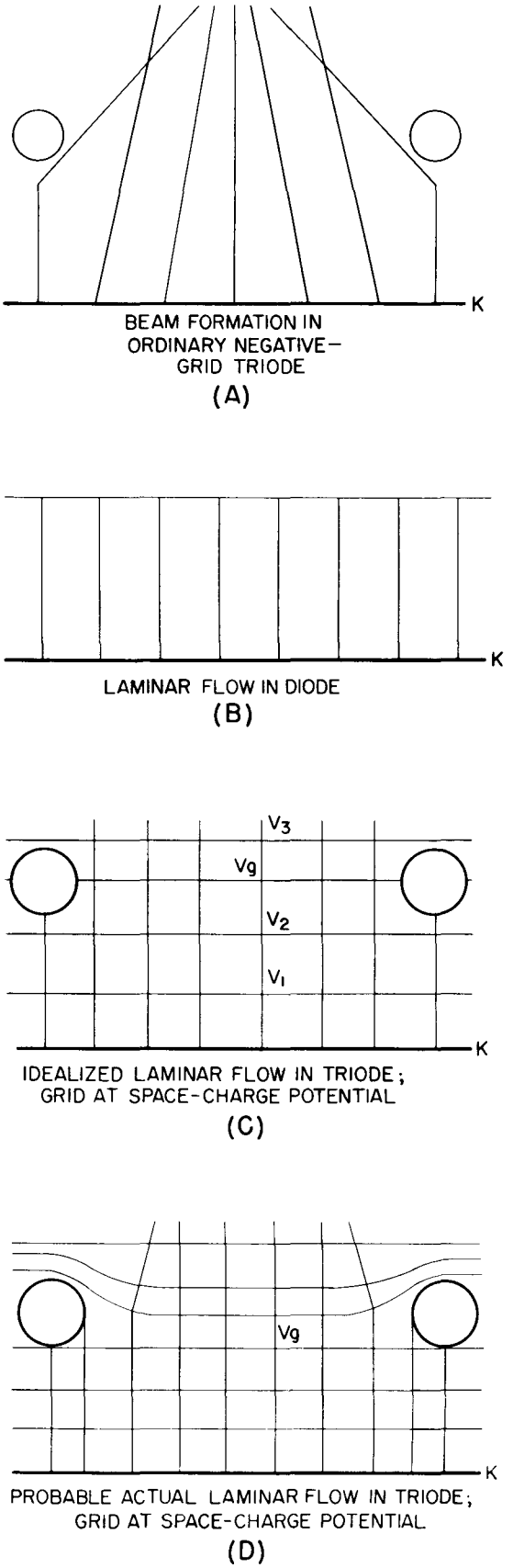


Figure 19. Electron Flow Patterns in Diodes and Triodes

*This procedure was first described by O. H. Schade, Sr.

**This statement is not strictly true. For any wire of finite radius, the equipotential lines must bend around the wires in the vicinity of the wires. Laminar flow is probably established when the field has the shape shown in Fig. 17D, rather than in Fig. 17C.

Procedure for Analysis*

(1) Read I_K and $\hat{g_m}$ as a function of V_g with the ratio I_g/I_K equal to $2r/a$. The term $\hat{gm}$ is defined as the transconductance read with the grid at space charge potential and should be read at the cathode lead to include all the current, $I_K = I_g + I_a$.

(2) At a constant bias (e.g., -0.6 to -1.0 volts) or constant grid current (e.g., 0.1 microamp) read gm as a function of I_b by varying E_b.

(3) Plot log I_K and $\hat{gm}$ as functions of V_g. On the same graph plot g as a function of Vg. The diode conductance g may be obtained by graphical differentiation of the I_K curve. On the same sheet plot gm as a function of I_b; use the I_K curve as an axis, as shown in Fig. 20. The diode conductance g is the theoretically obtainable conductance for the tube structure under analysis. The term $\hat{gm}$, the maximum transconductance obtainable for the triode structure, is less than g by the factor of the denominator of Eq. (80). The term gm, the actual transconductance available, which is less than $\hat{gm}$ because of field nonuniformities may be increased (theoretically up to the value of $\hat{gm}$) by using more ideal grids. The break point of the I_K curve is used to determine I_∞.

The design of triodes is similar to the design of diodes except that the design G must be greater than the desired gm by two factors: (1) the denominator of Eq. (80) and (2) a factor of 2 or 3 for field roughness.

The capacitances may be estimated from field plots for the active portion of the structure. The total capacitance is the sum of these estimates plus the capacitances of the stem, support members, etc. The optimum triodes require the finest wire and highest pitch grids commensurate with the attainable interelectrode spacings, manufacturing techniques, and tolerances. The capacitances of the nonactive structure must be kept to a minimum.

TETRODES AND PENTODES

This section discusses multigrid tubes, especially the tetrode and pentode; the analysis may be extended to other multigrid tubes.[28, 29, 30, 34] In this basic approach, equivalent potentials for the various grids are calculated, and ultimately, the structure is reduced to an equivalent diode. Because of the additional electrodes at positive potentials, an additional factor, current distribution, must be examined. It has been shown that the space charge has a marked effect on the current distribution, plate resistance, and overall simplification factor.

*This analysis is an attempt to provide a more exact and scientific means of evaluating tubes, but does not include a good deal of subsequent work. The author suggests referring to Dingwall[67] and the article in this book by O. H. Schade, Sr. for a description of the latest work on this subject.

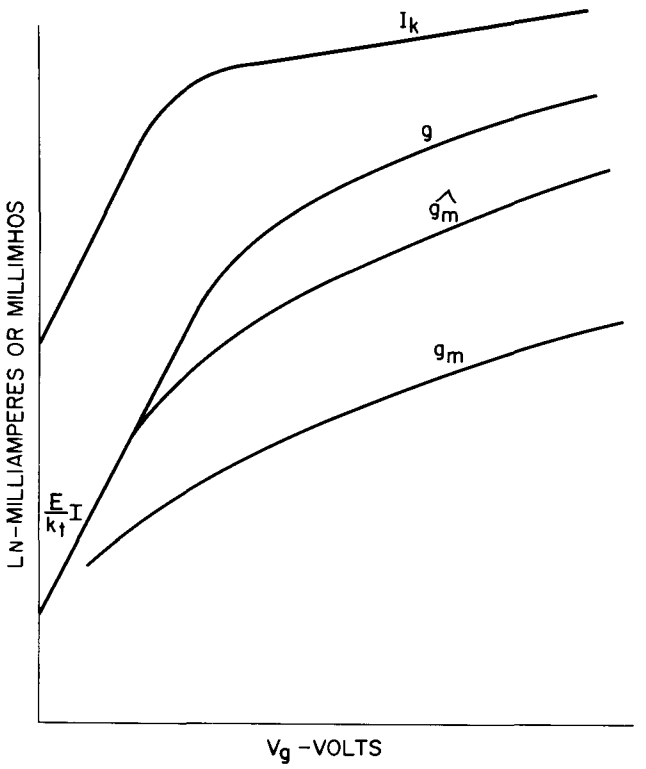


Figure 20. Typical Plot g, $\hat{gm}$, gm, and I_K vs. V_g

The tetrode, which was originally designed to reduce the grid-anode capacitance of the triode, has found wide use because of the form of the plate current-anode voltage characteristic. Although the triode is limited in that the voltage available for accelerating the electrons varies with the plate voltage swing, the tetrode has a fixed voltage available (the screen-grid voltage) which makes it especially useful when power output or a large voltage swing is required. However, the tetrode still finds widest use when low grid-anode capacitance is a requirement.

Early tetrodes were limited by the characteristic dip resulting from secondary emission from the plate. In the pentode, a third grid was introduced to suppress these secondary emissions. However, other means of suppression are used, and many tetrodes are still designed. The beam power tetrode, an important subgroup, is discussed separately in the article, "Beam Power Tube Design Considerations" by O. H. Shade, Jr., in this book. The design of pentodes and tetrodes depends mainly on the proper designing of the suppression mechanism (sometimes referred to as back-end design). The front end (cathode, grid No. 1, and grid No. 2) may be designed using triode principles. Many modern pentodes for wide-band amplifier service are essentially tetrodes which use a suppressor grid or beam plate only to reduce interelectrode capacitances.

Equivalent Potentials

The equivalent potentials for the pentode V_{eqv} (see Fig. 21) are derived from the basic equation

$$V_{eqv} = \frac{V_1 D_{12} + V_2 + V_3 D_{32}}{D_{12} + 1 + D_{32}}$$

From this equation, the equivalent potentials for each of the grids may be written directly as follows:

$$V_1 = \frac{D_{21}\ V_2 + V_g}{1 + D_{21} + D_{0i}} \qquad (81)$$

$$V_2 = \frac{D_{12}\ V_1 + V_s + D_{32}\ V_3}{1 + D_{12} + D_{32}} \qquad (82)$$

$$V_3 = \frac{D_{23}\ V_2 + V_{sp} + D_{43}\ V_p}{1 + D_{23} + D_{43}} \qquad (83)$$

D_{ij} is the penetration factor element i to element j.

V_i is the equivalent voltage of element i (numbered outward from K; K = 0)

where

V_p = plate voltage (applied)

V_g = grid voltage (applied)

V_s = screen voltage (applied)

V_{sp} = suppressor voltage (applied)

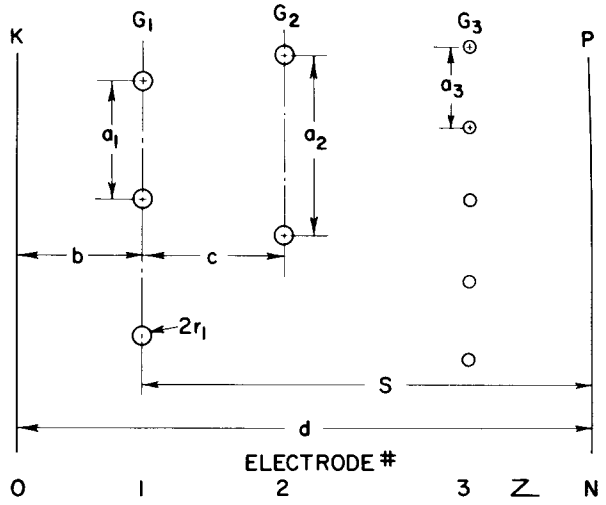
These equations could be combined to give a single expression for V_1. A more sensible approach is to eliminate the negligible factors and compute V_2 and V_3 and then V_1. These equations may be used for tetrodes by setting V_3 equal to V_p. The following procedure may be used:

(1) Calculate V_3, let $V_2 = V_s$.

(2) Calculate V_2, let $V_1 = 0$; use V_3 as calculated above.

(3) Calculate V_1; use V_2 as calculated above.

In the tetrode, the situation is more complex. There is generally a potential minimum between screen and plate, but the calculation of its position and magnitude is among the more difficult problems. In beam power tubes, it is generally assumed that the potential in the plane of the beam plate is zero. Just how important these inaccuracies are is a matter of conjecture. Usually the results will be close enough for the first trial and a small amount of empirical work will finish the design. When the equivalent diode has been determined, the cathode current and triode gm may be calculated. Because not all of the current reaches the plate, the front end must be over-designed. Because the screen is a grid, the field is not uniform in the plane of the screen. This condition contributes to the degradation of the theoretical transconductance ($\hat{gm}$) in the same manner as control grid field roughness. When the cathode current and triode gm have been calculated, the ratio I_a/I_s, the plate resistance rp, and the knee voltage (and suppression which is intimately connected with the knee voltage) may be determined as discussed below.



a_N = PITCH OF N^{TH} ELECTRODE
r_N = RADIUS OF GRID WIRE OF N^{TH} GRID
b = GRID 1 – CATHODE SPACING
S = GRID 1 – ANODE SPACING
c = GRID 1 – GRID 1 SPACING
d = CATHODE – ANODE SPACING

Figure 21. Cross Section of Multigrid Tube Giving Nomenclature for the Various Spacings

The Effect of Space Charge in the Anode Region

Because a mathematical analysis of the space-charge effects has been described elsewhere and is somewhat involved, this section is restricted to a discussion of some of the ramifications of the analysis.

Qualitative Discussion. It is assumed that the initial velocity for the emitted electrons is zero and that any point in the beam up to the screen grid has a single velocity. The electrons are assumed to travel in straight parallel paths as they approach the screen. These assumptions are also generally used in mathematical treatments. Based on the above assumptions, as the beam passes through the screen, a small portion of the current is directly intercepted, and the remainder continues on towards the plate. If space charge is neglected (or is very small) the electrons reach the plate even if the plate potential is zero because they originated at zero potential. However, we know that for reasonable current densities the space charge becomes important, and the potential variation is such that, for low plate voltages, some of the current is turned back (thermal velocity spread must then be considered). There is a critical value of plate voltage, the knee voltage, above which all the injected current is collected. The knee voltage is not generally precise or sharply defined because the electrons passing near the wires of the screen grid are deflected laterally. This effect gives the beam a lateral spread and introduces a velocity spread. The electrons having the greatest entrance angles require the highest voltages to be collected and, as a result, the knee is rounded. In an actual tube, the electrons approaching the screen usually have a considerable angular spread which causes further rounding of the knee.

Mathematical Solution-References. The mathematical analysis of the problem has been carried out in great detail by Fay, et al[5] and Salzberg and Haeff.[4] Both papers give graphical data that may be used to compute the form and magnitude of the potential distribution. Such factors as the knee voltage, the voltage available for suppression, and maximum current may be calculated. Such anomalies must be expected; the mathematics is not always single-valued and hysterisis effects have been predicted and observed. The geometry in a real tube deviates widely from the ideal and this, of course, gives additional errors.

Screen-Grid Current. The calculation of screen-grid current is always difficult. The directly intercepted current may be estimated, but the current returned from the plate space may make many excursions through the screen plane before it is collected. This makes accurate calculation of screen current difficult. Some work has been done on this problem.[31, 32, 33, 35] Low screen current is important in power output tubes to provide high efficiency and safe screen temperatures. In rf amplifier tubes, screen current causes noise (partition noise) and low screen current is desirable. For power output tubes, the control and screen grids are aligned to provide focusing. Because wide-band amplifiers have low-pitch control grids and higher values of mu, it is not generally possible to line up the grids. Low screen current may be achieved by careful design of the back end or by special construction (i. e., shielding the screen grid with a third grid run at zero potential). This has been used in the past, but generally requires high voltages or complicated structures and is only economical in extraordinary circumstances.

Plate Resistance.[36] Very little can be said about this calculation of plate resistance, other than that it is largely controlled by the space charge in the anode region. We know that the mu is affected by space charge (the cathode space in the triode analysis), but there are also further complications such as secondary emission and reflected electrons returned to the screen grid.

VARIABLE MU EFFECTS

The Concept of Variable Mu

The preceding sections considered only ideal tube structures; this section discusses some deviations from the ideal. The ideal triode is assumed to have a perfect grid that controls the field at the cathode without intercepting or deflecting the electrons. The electric field intensity at the cathode will then be a constant along the cathode in either of the two dimensions. Such an ideal structure has a constant mu that is not attainable in practice. Nonconstant or variable mu introduces nonlinearity into the characteristics of the tube (over and above the basic nonlinear 3/2-power relation). This additional variation may or may not be desirable. For radio and television systems, a wide range of signal amplitudes is handled without distortion by tubes having a remote cutoff characteristic. On the other hand, because variable-mu tubes suffer some degradation in gm and performance, it is desirable to eliminate any unwanted variation to maximize gain.

Variable mu effects may be categorized as (1) designed effects, which result from variable-pitch grids and other methods, and (2) accidental effects, which result from nonideal grids, variable spacing, or statistical variations (in electrode dimensions). This section describes an approximate method of design for variable pitch grids and discusses the accidental variations.

Design of Variable-Mu Tubes

A mu variation is used to extend the grid base (remote cutoff) and provide large signal handling capacity.

In an ideal tube, having a grid voltage for cutoff of E_{bb}/μ, the cutoff may be extended by winding the grid so that the main section has low pitch and high mu, and one or more sections of lower mu. The various sections of the grid cut off at different values of grid voltage, and, at low bias, the characteristic is a composite (see Fig. 22). If the characteristics of the individual sections could be computed, the overall characteristics could be drawn. Fig. 22 shows idealized linear approximations; before attempting to calculate actual characteristics, some general aspects of the problem will be discussed.

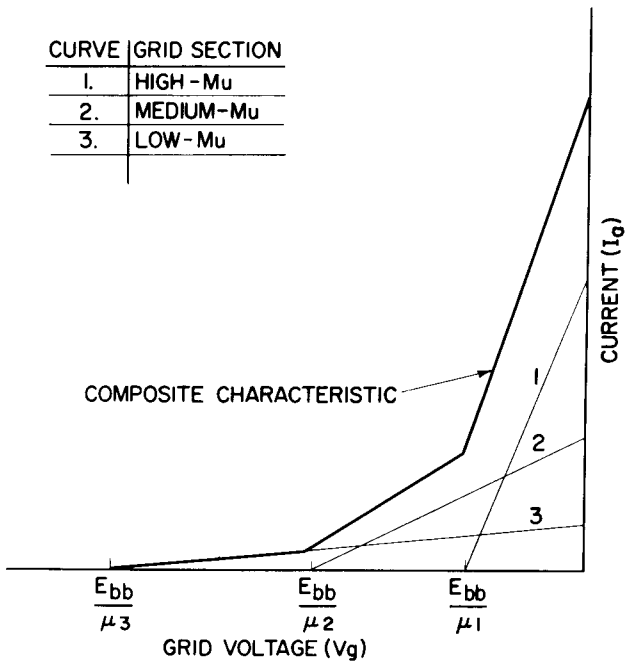


Figure 22. Composite Transfer Characteristic for Variable Mu Tube

The lowest value of mu is governed by the cutoff bias desired for the tube. The area of this section (i. e., number of turns) is, in turn, governed by the gm and/or current desired at some bias less than the cutoff value. The intermediate sections are determined by the curvature desired in the transfer characteristic. The main section is determined by the gm and current required near zero bias. In practice, the various sections of the grid are chosen and the composite curve computed. The design is varied until the desired composite curve is achieved.

This section describes how the contribution of the individual sections may be combined to give the composite characteristic. It should be realized that the following discussion is general and does not refer to variable-pitch-grid tubes only, but to any combination of nonuniform tubes. In fact, it applies to the present case only insofar as the sections are separate and distinct. Any interaction between sections would somewhat invalidate these results.*

The total plate current I is the sum of the currents through each section of the grid.

$$I = i_1 + i_2 + i_3 + \cdots + i_n \tag{84}$$

Differentiation with respect to V_g gives

$$\frac{\delta I}{\delta V_g} = \frac{\delta i_1}{\delta V_g} + \frac{\delta i_2}{\delta V_g} + \frac{\delta i_3}{\delta V_g} + \cdots + \frac{\delta i_n}{\delta V_g} \tag{85}$$

or

$$Gm = g1m + g2m + g3m + \cdots + gnm \tag{85a}$$

Differentiation with respect to V_a gives

$$\frac{\delta I}{\delta V_a} = \frac{\delta i_1}{\delta V_a} + \frac{\delta i_2}{\delta V_a} + \frac{\delta i_3}{\delta V_a} + \cdots + \frac{\delta i_n}{\delta V_a} \tag{86}$$

or

$$Ga = g1a + g2a + g3a + \cdots + gna \tag{86a}$$

The composite transconductance and plate conductance are the sums of the individual conductances because the tubes are in parallel. The composite mu is determined from the following expression for current:[67]

$$I = \frac{KA}{B^2}\,(DV_a + V_g)^{3/2} \cdot F \tag{87}$$

$$B = b - dm,\quad V_a = V_a - V_m - \phi_K + \phi_a$$

$$F = \phi(\eta^+),\quad V_g = V_g - V_m - \phi_K + \phi_g$$

If this expression holds, with the appropriate values, for both the composite tube and the individual sections, the expression may be written as follows:

$$\frac{KA}{B^2}(DV_a + V_g)^{3/2} F = \sum_{i=1}^{N} \frac{KA_1}{Bi^2}(D_i V_a + V_g)^{3/2} F_i \tag{88}$$

The values for the composite tube are average values, as discussed below. This equation may be solved for the composite Durchgriff D as follows:

$$D = \frac{1}{V_a}\left\{\left[\sum_1^N \frac{a_i}{b_i^2} f_i (D_i V_a + V_g)^{3/2}\right]^{2/3} - V_g\right\} \tag{89}$$

where $a_i = \frac{A_i}{A}$, $b_i^2 = \frac{B_i^2}{B^2}$, $f_i = \frac{F_i}{F}$

Some approximations must be made when applying these results.

The grid-cathode spacing may vary if the actual geometry varies or if the potential minimum position varies because of varying current density. The average value used to compute the composite current is taken as the mean-square average of the individual spacings.

$$B = \sqrt{\frac{\sum B_i^2}{\eta}}$$

The values of F, V_m, ϕ_K, ϕ_a, and ϕ_g should be average values for the composite structure. Eq. (89) may be used to compute the Durchgriff for variable pitch grid, variable spacing, or even varying work functions. For a quick check on variable-pitch tubes, the refinements may be dropped and the simpler Child's formula may be used as in the following example of the 6FW8 variable-pitch grid tube shown in Table I.

Mu Variations from Nonideal Grids

There are many mathematical formulations for the mu of a helical grid between two plane electrodes. This calculated mu and its subsequent application in the calculation of equivalent diodes is usually limited to the case where $b > a$ and $r << a$. Very few modern tubes meet these criteria and most are characterized by the existence of a phenomenon known as inselbildung or shadow effect (see Fig. 23). Under these conditions, the field strength and current density at the cathode surface are not constant. Depending upon the operating condition, some areas of the cathode may be cut off before others. Measurement of mu as a function of grid voltage shows a decrease with increasing bias, as shown in the published data sheets for any modern tube. This variation has two effects which must be considered:

(1) the normal tailing-off of the transfer curve and the reduction in gm resulting from nonuniformity

(2) cathode current may be less than predicted from theoretical mu values, because some areas of the cathode may be cut off.

An estimate of these effects requires some idea of the magnitude of the variation of mu and current density. This variation has been determined empirically and mathematically by Fremlin,[68,69] mathematically by Bennett and Peterson,[37] and Dahlke,[38] on the analog field plotter by Napiorski, Schade, and Scholz,[39] and numerically by Harris and Kirk[40] on a digital computer. The data are conflicting even for the simple cases considered. The actual situation is greatly modifed by the space charge, and, to our knowledge,

*The mu of individual sections of a variable pitch grid may be difficult to determine accurately. The formulas presented apply to an infinite array of wires equally spaced. For short sections of the grid where the wire spacing changes abruptly the calculation is likely to be in error.

Table I

Calculation for Mu of 6FW8

Section	Length	Pitch	r/a	Da	ai	
1	0.2056	0.0022	0.091	0.025	0.745	S = 0.008
2	0.0500	0.0025	0.080	0.032	0.181	D_{avg} = 0.0287
3	0.0168	0.0028	0.071	0.042	0.061	
4	0.0032	0.0032	0.062	0.059	0.016	Mu = 34.8

The computed average of 34.8 compares favorably with the measured value of 33.

the complete solution has not yet been obtained. The results obtained by Bennett and Peterson, shown in Fig. 24, are valid only for small grid wires. The results obtained by Napiorski, et al are presented for comparison (Figs. 25 thru 28). Any work of this sort must be considered as a first approximation and interpreted in the light of the assumptions that were made in its derivation. The numerical solutions[40] are probably the most accurate, but have not yet been made generally available; a portion of this data is included in Fig. 29.

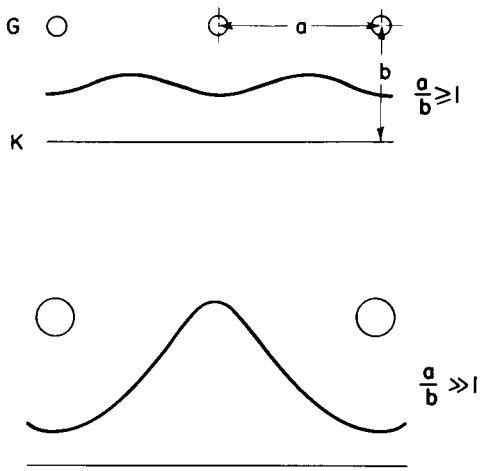


Figure 23. Current Flow Patterns Showing Inselbildung

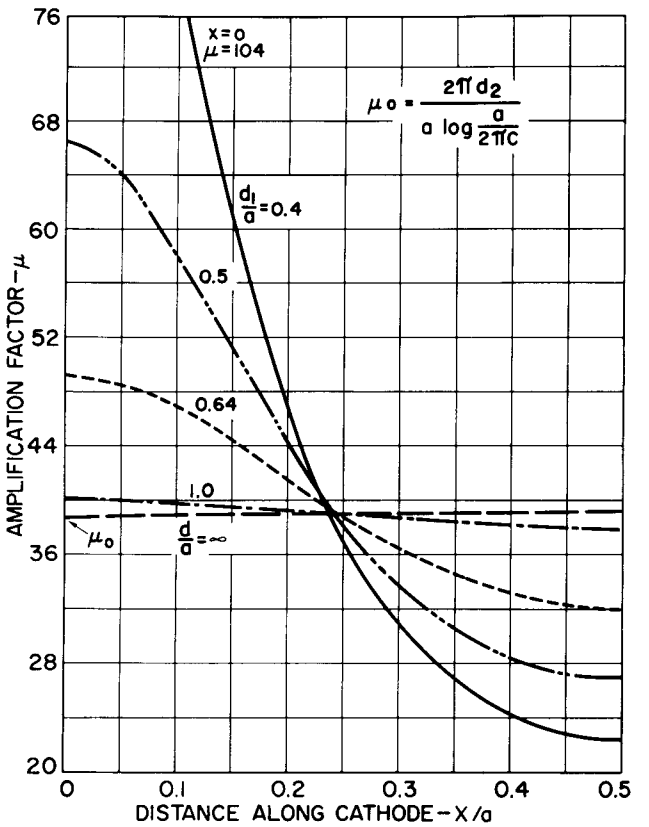


Figure 24. Inselbildung Effects in Triodes - Bennett and Peterson: (A) Variation of mu along cathode surface of triode;

Variable Spacing and Statistical Variations

The problem of variable or nonuniform spacing may be handled in the same manner as the variable-pitch grid, but requires more accurate analysis to obtain useful results. There is evidence that for variations in grid-cathode spacing inselbildung effects may outweigh other effects and make analysis difficult.

Statistical variations in grid pitch or tube interelectrode spacing also lead to variable mu and degradation of tube performance. The difference between statistical variations and shifted parts is one of degree and may be handled in a similar manner except that statistical methods must be used. This problem has been treated theoretically by Dahlke.[41]

FIELD PLOTTING

The previous sections derived current-voltage relations for common geometries without explicit calculation of the electric fields. However, it is sometimes necessary or desirable to actually plot the field for electron trajectory problems, estimation of capacitance or the determination of points of high field for voltage-breakdown consideration.

Mathematical Plotting Methods

Direct Solution of the LaPlace Equation. Direct solution of the LaPlace equation is possible only for simple geometrical configurations, and, therefore, the direct application of this method to electron tubes

is limited. However, small sections of the complex tube structure may be successfully investigated by this method.[42]

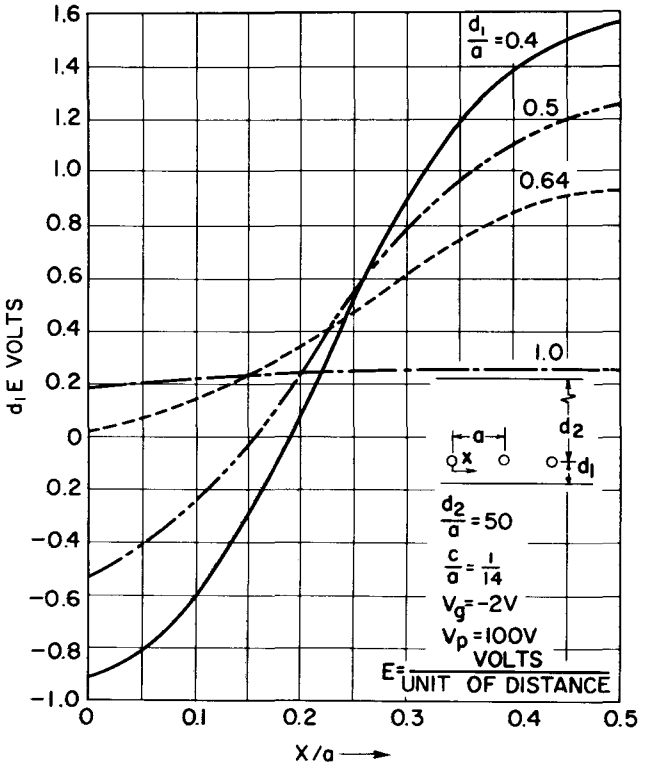


Figure 24(B). Variation of field strength along cathode surface of triode

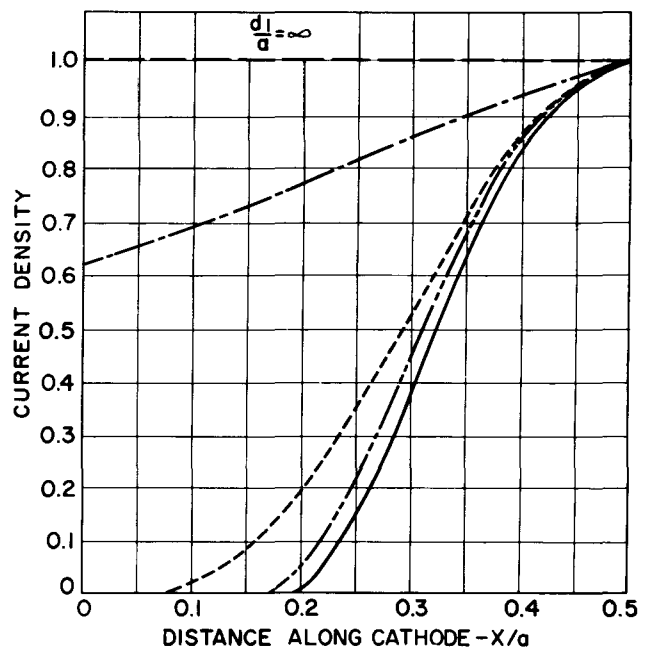


Figure 24(C). Variation of current density along cathode surface of triode

Conformal Mapping. In a conformal map, orthogonal functions retain their orthogonality when mapped.* As a result, it is possible to map a complex geometry and transform (warp) it into a simpler one for which a solution is obtainable. This method has many applications and has often been used (e.g., Vodges and Elder[21] and Herne[22]). A collection of conformal mapping functions have been made by Kober[43] and Breckenback.[44]

Numerical Solutions. Many highly sophisticated numerical methods for the solution of partial differential equations may be found in the literature.[45, 46] However, some relatively simple approximation methods are also available. Consider a subdivided region having an electric field as a function of the potential; if the divisions are fine enough, a linear approximation of the derivatives may be used, as follows. (See Fig. 30.)

$$\left.\frac{\delta V_a}{\delta x}\right|_a = \frac{V_i - V_o}{\delta} \quad \& \quad \left.\frac{\delta V_b}{\delta x}\right|_b = \frac{V_o - V_2}{\delta} \qquad (90)$$

In the center at point 0, the second derivatives are evaluated and the substitution is made in the LaPlace equation.

$$\frac{\delta^2 V_o}{\delta x^2} + \frac{\delta^2 V_o}{\delta y^2} = 0$$

$$\frac{\delta^2 V_o}{\delta x^2} = \left(\frac{\delta V_a}{\delta x} - \frac{\delta V_b}{\delta x}\right)\frac{1}{\delta} = \frac{1}{\delta^2}(V_1 + V_2 - 2V_o) \qquad (91)$$

A similar result is found for $\delta^2 V_o/\delta y^2$ and LaPlace's equation is found to be equivalent to $V_o = 1/4\,(V_1 + V_2 + V_3 + V_4)$.

For small subdivisions, the center potential is just the average of the four surrounding potentials. This relationship is the basis of the relaxation net or point net methods in which the region around a set of electrodes is subdivided and potentials are assigned arbitrarily to the intersections. The assigning of the potentials makes full use of known boundaries and symmetry planes.

The assigned potentials are used to calculate a new set of potentials which, in turn, is used to recalculate the initial values. The process is repeated until the values show a minimum change with each new determination. The solution of the problem is facilitated if values that show excessive change are adjusted before recalculating. Southwell has formalized a somewhat modified procedure.[47, 48] (See Fig. 31.)

Analog Plotting Methods[49]

There are three well-known analogs: (1) the resistance paper or electrolytic tank,[50, 51] (2) the rubber sheet,[52, 56] and (3) the resistor net.[53, 54, 55] In each of these systems, two quantities may be found that are related, in the first order, by the LaPlace equation. These quantities are generally easier to measure than

*Equipotential and flux lines retain their relationship.

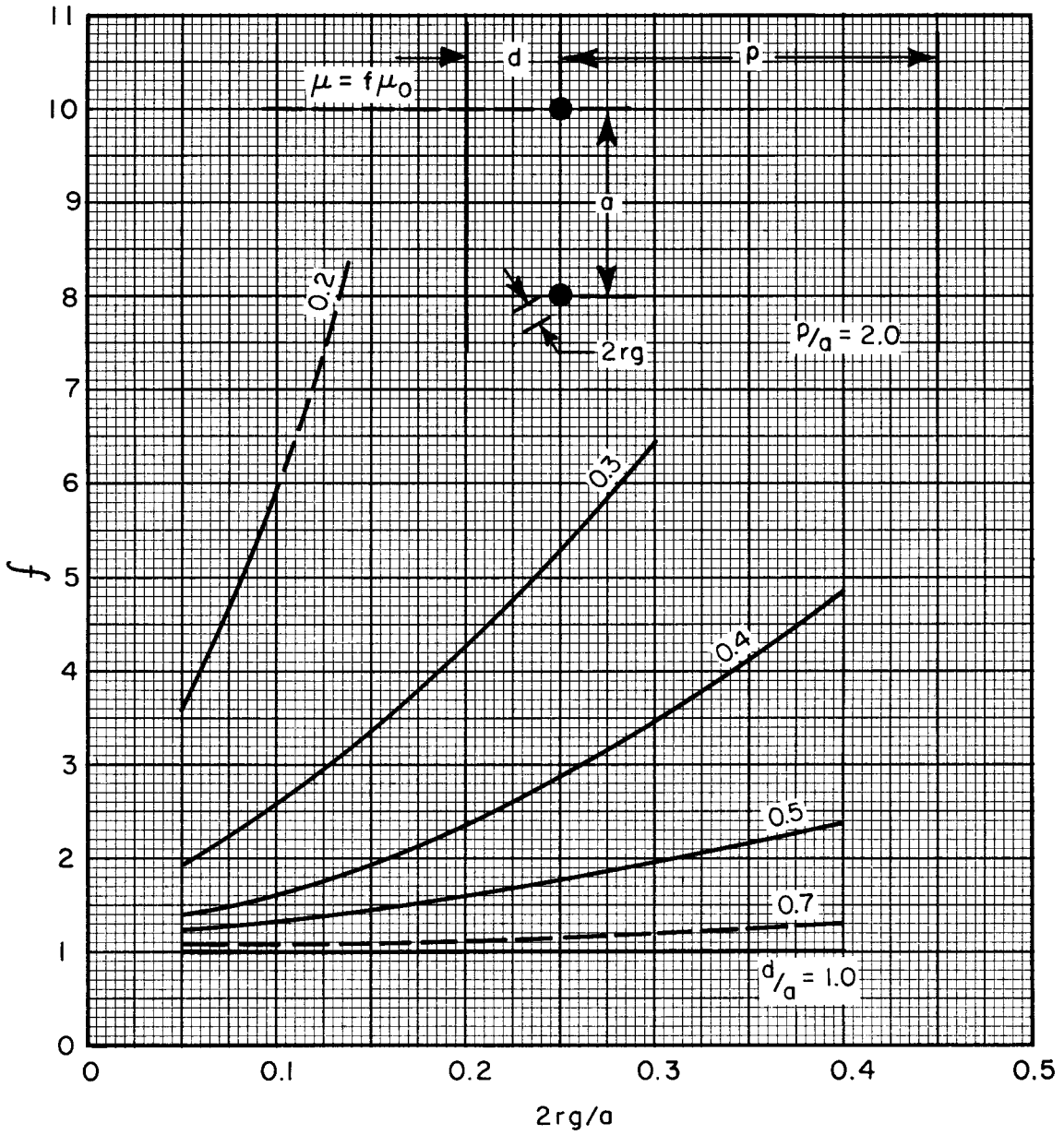


Figure 25. Increase of Mu Directly behind a Grid Wire

electrostatic flux and/or potential. The rubber sheet analog has been little used in recent years and is not discussed here.

The Resistance-Paper or Electrolytic-Tank Method. In these analogs, flux is equated to current and potential difference is equated to voltage difference. A model of the device to be mapped is made on either special conducting paper or in an electrolytic tank (the paper and the tank may be considered the same except that the tank is sometimes more versatile). Maximum use must be made of the boundaries and symmetry to simplify the model. The problem is considered in only two dimensions. The actual device is reduced to two dimensions either by assumption or by using sections and making a series of plots. The basic concepts of field plotting are shown in Figs. 32 thru 38.[57]

The following conditions must be considered when making a field plot:

(1) Lines of symmetry are flux lines in an actual field.
(2) Current flow must follow the edges of the paper, therefore, the edge of the paper is a flux line. (The paper may be cut to a special shape.)
(3) If the edges of the paper are not used as a boundary, the plot must be kept away from the edges to limit the perturbation of the field.
(4) Periodic structures may be represented by a single section by using the edges of the paper as a "reflecting" boundary.[20, 51]
(5) Sections of a tube in which the grid wires represent circles present no special problems, but sec-

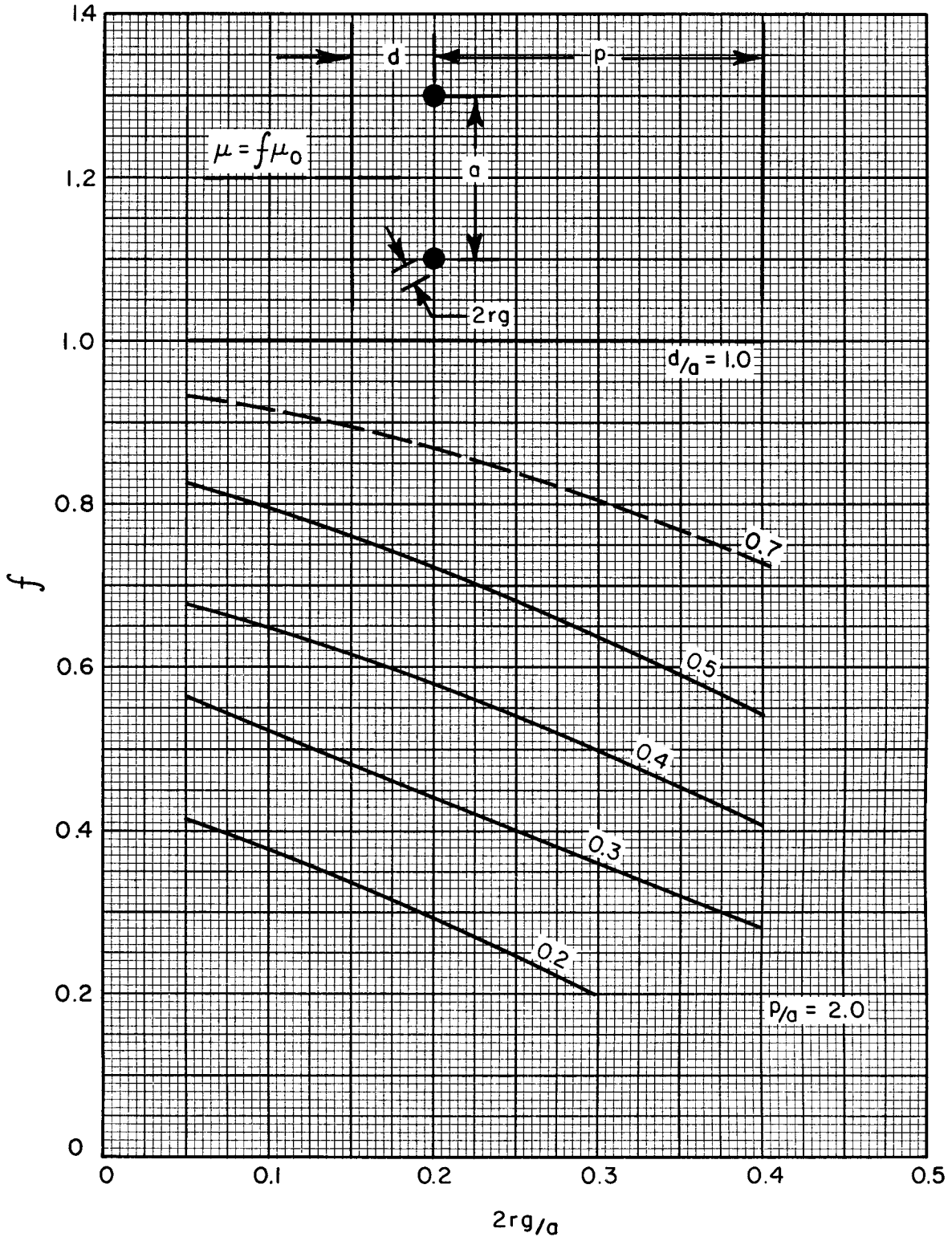


Figure 26. Decrease of Mu at the Center of the Aperture

tions in the other direction may be difficult to represent, for you must either go through a grid wire or between the wires. Either choice is essentially wrong. It is sometimes better to establish an equivalent potential in that plane by means of a fine-wire low-pitch grid of equivalent mu, as shown in Fig. 39.

(6) Space charge is normally neglected.

(7) Before making a plot, decide exactly what it is you will learn from the plot.

(8) The accuracy of the plot is affected by the variations of the paper, the accuracy of the model, and the degree to which measurements can be made without

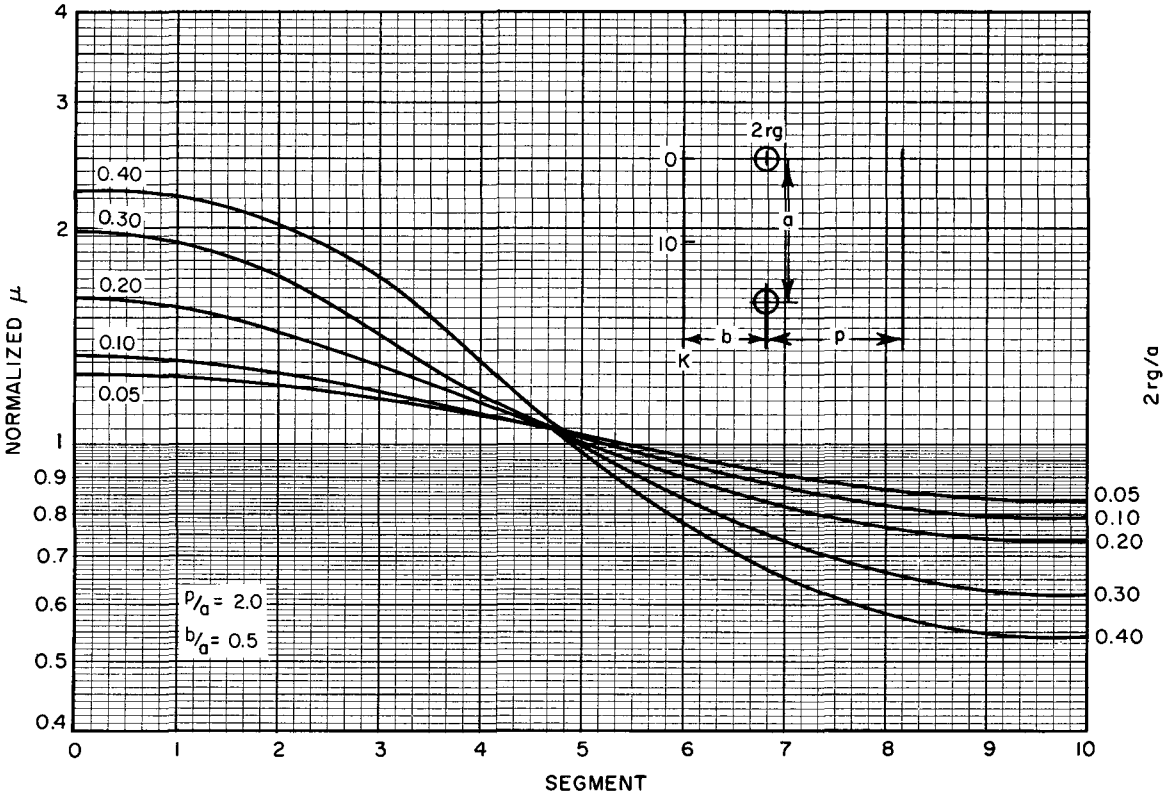


Figure 27. Mu Variation along Cathode 2r/a as a Parameter (Scholz, et al data)

disturbing the current. The electrolytic tank probably has a greater accuracy.[58]

(9) By tilting the electrolytic tank, one may represent a wedge of a cylinder or various squares or polygons.[20, 51]

(10) Methods are available for simulating space charge by injecting current into the tank or resistor wedge.[59] This technique does not apply to the paper method.

(11) Theoretically, the scale has no effect on the solution, but, in general, the larger the scale the greater the relative precision possible in the model.

(12) It is possible to make three-dimensional models in the tank and, by probing at various depths with an insulated fine probe, to make a three-dimensional plot.

The Resistor Net. The resistor-net method is essentially a lumped parameter or discrete approximation of the paper or tank, representing either rectangular or cylindrical geometry. Space charge may be represented by successive approximations. The data is available in digital form and lends itself to automatic plotting of trajectories.[60, 65] This method is especially well-suited to cylindrical beam problems.

Field Sketching

This method, although it gives only approximate results, requires a firm understanding of the basic relations of electrostatics. A sketch is made of the probable field configuration based on simple structures and the basic relations, as shown in Fig. 40. A curvilinear square is formed by the intersection of four curved lines; the intersections are perpendicular and the figure tends to become a square as it is subdivided. Sketching flux and potential lines and forming curvilinear squares results in an acceptable plot from which trajectories, gradients, and capacitances can be estimated. Capacitance may be estimated as shown in Fig. 40. The capacity C is, by definition, equal to ϕ/V, where ϕ is the total flux and V is the potential difference. For a square of width d, the gradient E is V/d and the flux density D is ϵE. Multiplication by the area provides the following relationship for a one-meter length:

$$\phi = \epsilon Ed = \frac{ev}{d} d \qquad (93)$$

and, therefore,

$$C = \epsilon \text{ farads/meter} \qquad (93a)$$

Because the voltage across a square and the flux through it is the same for all squares, we find the total capacitance to be

$$C = \frac{P}{Q} \epsilon \text{ farads/meter} \qquad (94)$$

where P is the number of squares around one of the electrodes and Q is the number between them. In the example of Fig. 40, P = 18, Q = 4, $\epsilon = 1/36\pi \times 10^{-9}$, and C = 29.8 picofarads/meter.

Electrode Fitting

Generally, the electrode configuration is given and

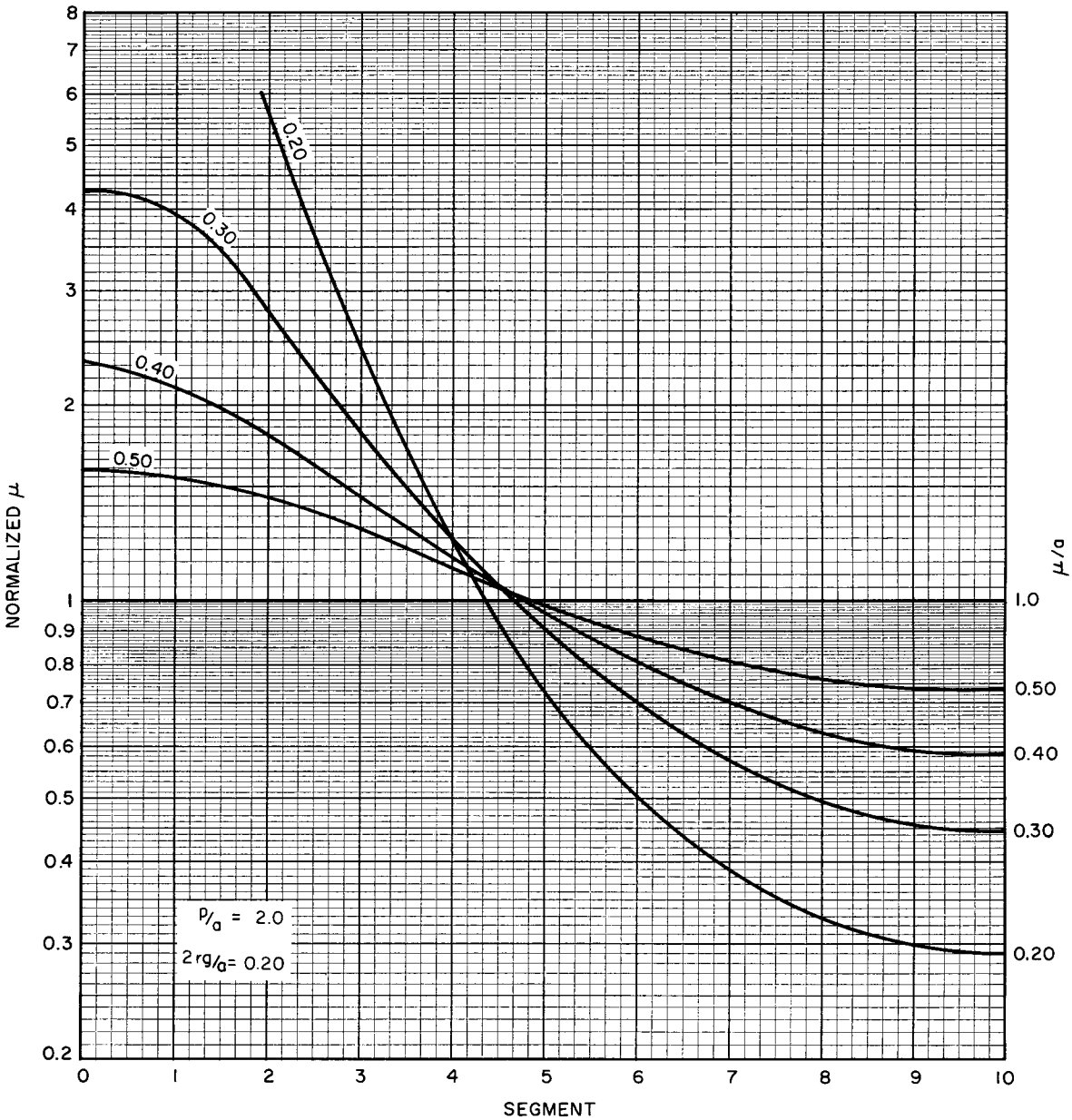


Figure 28. Mu Variation along Cathode b/a as a Parameter (Scholz, et al data)

the field distribution must be determined. In electron beam devices, the beam boundaries are usually specified, the field required to hold the beam within these boundaries is calculated, and then a set of electrodes and potentials are designed which provide the required field. The field requirements are generally in the form of required potentials along the edge of the beam. The determination of these potentials is exceedingly difficult. Although there is only one field configuration for each set of electrodes, and only one field configuration that will give the proper beam, many sets of electrodes will give approximately the proper field along the beam edge. Therefore, the set of electrodes is most easily found empirically (e.g., by an analog plot). This method simulates the edge of the required beam by shaping the paper or providing a dielectric boundary in the tank. A set of electrodes is tried and then the shape or relative potentials are adjusted until a fit is reached. This technique can be made almost automatic by switching back and forth between the field map and a reference voltage and adjusting the electrodes for minimum error.[61, 62]

TRAJECTORY TRACING

General Equations of Motion[64]

In an electrostatic field, the force on an electron is represented as follows:

$$F = -e\underset{\sim}{E} = e\ \nabla V \tag{95}$$

and the general equation of motion is

$$M\underset{\sim}{S} = \nabla V \tag{96}$$

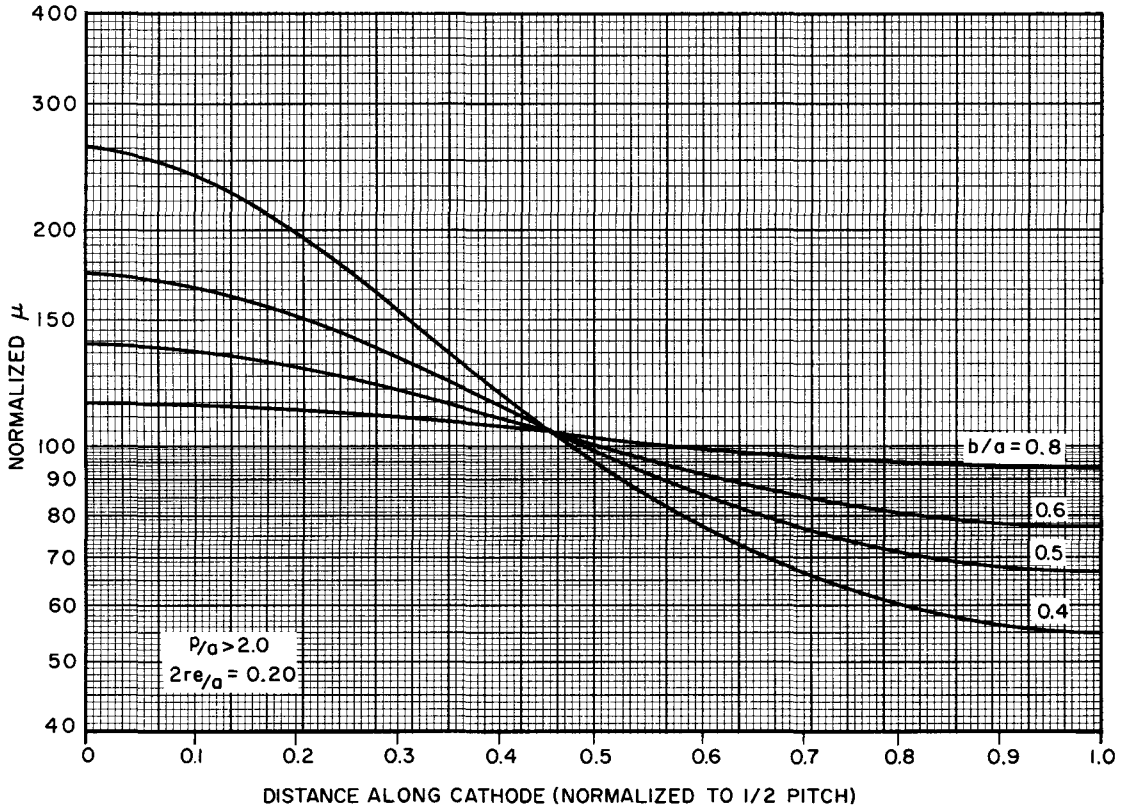


Figure 29. Mu Variation along Cathode (Harris-Kirk data)

In rectangular coordinates:

$$m\ddot{x} = e\,\frac{\delta V}{\delta x} \tag{97a}$$

$$m\ddot{y} = e\,\frac{\delta V}{\delta y} \tag{97b}$$

$$m\ddot{z} = e\,\frac{\delta V}{\delta z} \tag{97c}$$

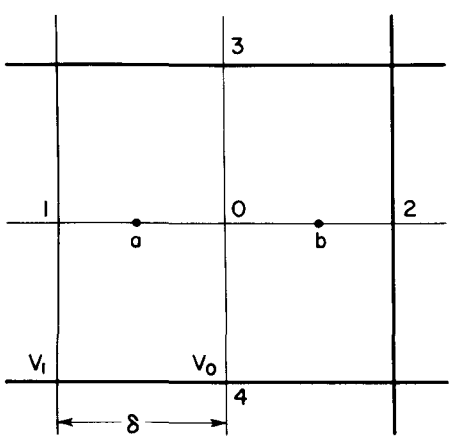


Figure 30. Subregion for Net-Point Analysis

These nonrelativistic equations apply as long as the mass of the electron is constant. The motional mass of an electron is related to the rest mass by the following transformation:

$$m = \frac{M_0}{\sqrt{1 - V^2/C^2}} \tag{98}$$

If the velocity is less than 0.1C (3×10^7 meters per second), the correction may be neglected. Because the electrostatic field is conservative, the following expression may be applied:

$$\frac{1}{2}mu^2 = eV \tag{99}$$

For an electron that starts at rest from a source at zero potential,

$$u = \sqrt{\frac{2e}{m}}\,V^{1/2} \tag{100}$$

or $u = 5.95 \times 10^5\ V^{1/2}$ meters per second, $u = 6 \times 10^7$ meters per second corresponds to a potential difference of approximately 10 kilovolts and is important only for kinescopes or special high-voltage types.

General Projectory in Two Dimensions

By eliminating time in the force equations it is possible to obtain the general trajectory equation.

$$2\,\frac{d^2y}{dx^2} = \left[Ex\,\frac{dy}{dx} - Ey\right]\left[1 + \left(\frac{dy}{dx}\right)^2\right] \tag{101}$$

This equation is not suitable for direct solution, but may be numerically integrated, especially if the potential is available as point-by-point numerical data.

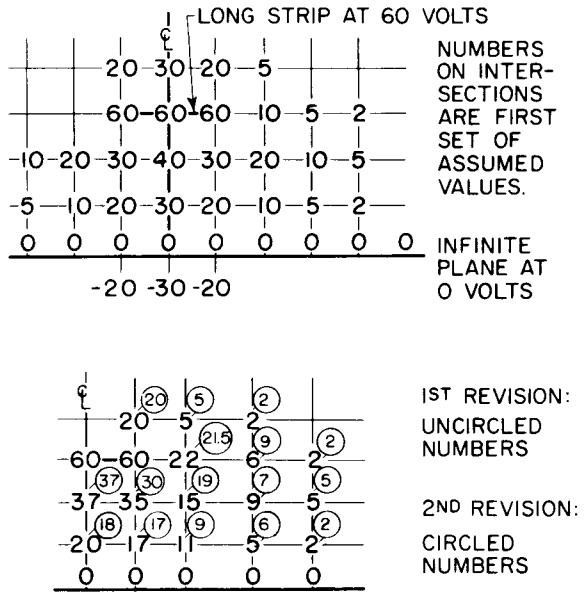

LONG STRIP AT 60 VOLTS
NUMBERS ON INTER-SECTIONS ARE FIRST SET OF ASSUMED VALUES.
INFINITE PLANE AT 0 VOLTS
1ST REVISION: UNCIRCLED NUMBERS
2ND REVISION: CIRCLED NUMBERS

Figure 31. Sample Problem Point-Net Method

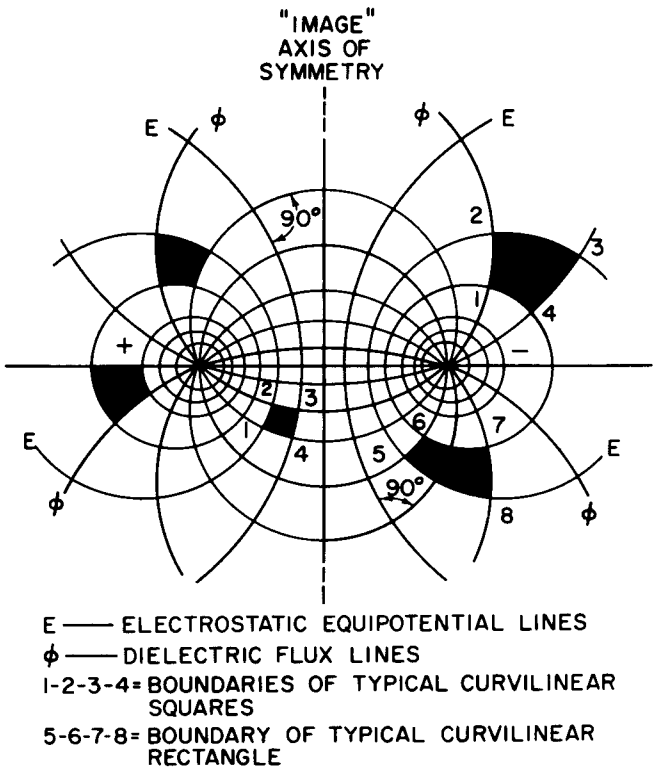

"IMAGE" AXIS OF SYMMETRY
E —— ELECTROSTATIC EQUIPOTENTIAL LINES
ϕ —— DIELECTRIC FLUX LINES
1-2-3-4= BOUNDARIES OF TYPICAL CURVILINEAR SQUARES
5-6-7-8= BOUNDARY OF TYPICAL CURVILINEAR RECTANGLE

Figure 32. Basic Orthogonal Field Concept

Solutions for specific electrode configurations are not generally available, but the following general conclusions drawn from the above result are given by Harman:[63]

(1) The path does not depend on the mass or charge and is the same for all particles.
(2) The velocity at which the particle moves depends on the charge-to-mass ratio.
(3) The path does not depend on the magnitude of the field (potential), but only on the configuration (gradient), and if the potential is everywhere multiplied by a factor K the path will be unchanged, but the time required to transverse the path is reduced by a factor 1/K.
(4) The path is unchanged if all the dimensions are scaled by a constant factor.
(5) The trajectory is not a function of the direction of travel.

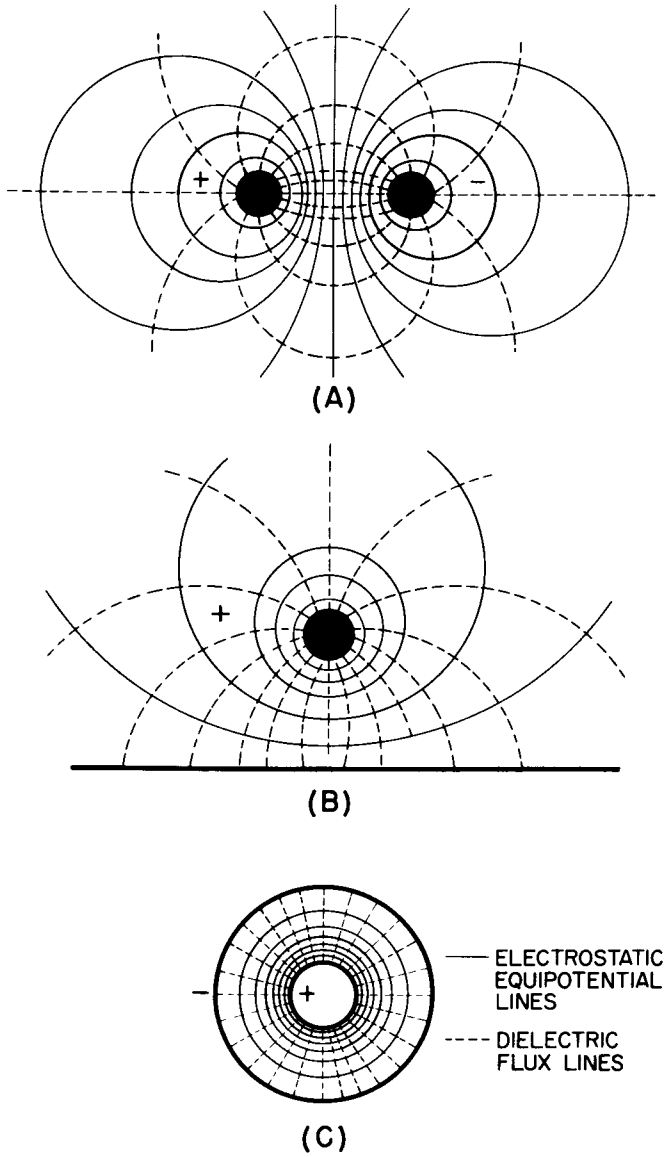


Figure 33. Fundamental Field Patterns: (A) Electrostatic field pattern between two oppositely charged parallel cylinders; (B) Electrostatic field pattern between oppositely charged plane and cylinder; (C) Electrostatic field pattern between oppositely charged concentric cylinders

Electron Motion in a Uniform Electric Field

In this simple case, the field is in only one direction y, the motion is described by the following equation:

$$M \frac{d^2y}{dt^2} = eEy$$

For the general case of an electron injected with a velocity U_o at an angle of σ with the field (see Fig. 41), the following expression is obtained for y (a parabola):

$$y = x \tan \sigma - x^2 \frac{E_y}{4V_o \cos^2 \sigma} \quad (103)$$

At a time expressed as

$$t_1 = \sqrt{\frac{2mV_o}{e}} \frac{\sin \sigma}{Ey} \quad (104)$$

the electron reaches its maximum penetration into the field.

$$y_m = \frac{V_o}{Ey} \sin^2 \sigma \quad (105)$$

and at a time $2t_1$ the electron has returned to the initial potential and crosses the x axis at the value

$$x_m = \frac{4V_o}{Ey} \sin \sigma \cos \sigma \quad (106)$$

By dividing the field into small sections, each essentially uniform, parabolic path segment may be calculated and the trajectory constructed. This is the basis for O. H. Schade's method with preplotted parabolas, explained below. Two methods of plotting trajectories are discussed below.

Snell's Law

Segmented Straight-Line Trajectories. Many of the laws of ordinary geometrical optics have analogies in electron optics.

$$\frac{\sin \alpha}{\sin \beta} = \frac{\eta_2}{\eta_1} \quad (107)$$

The familiar Snell's law for a light wave traversing the boundary between two media can be extended to electronics for an electron traversing the boundary between two uniform electrostatic fields.

$$\frac{\sin \alpha}{\sin \beta} = \frac{u_2}{u_1} = \sqrt{\frac{V_2}{V_1}} \quad (108)$$

Where α and β are the entrant and emergent angles, and η_1 and η_2 are the indices of refraction; the other symbols have the standard meaning. The application of this law is straightforward and simple.

Mechanics of Trajectory Tracing. The equipotential lines must be plotted at an interval small enough to allow the assumption of uniform fields between the equipotential lines. The trajectory is constructed of straight line segments as follows (see Fig. 42).

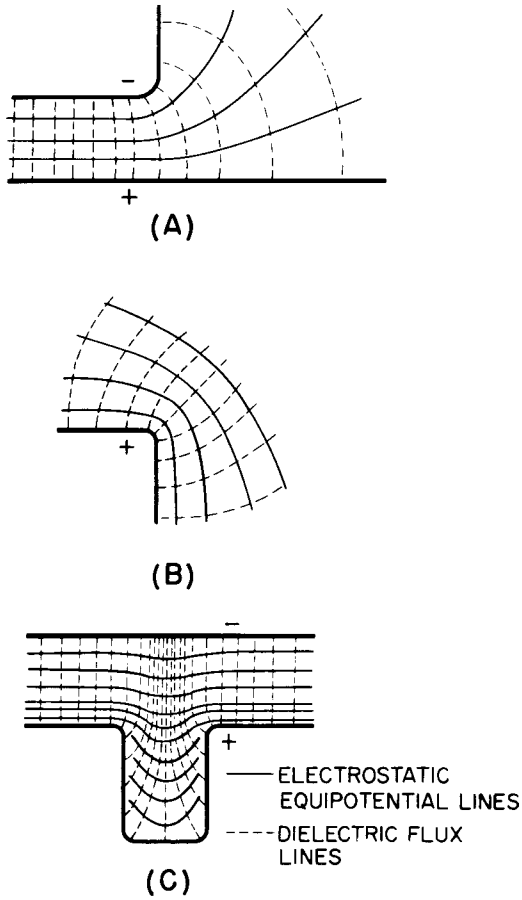


Figure 34. Typical Field Patterns: (A) Electrostatic field between oppositely charged corner and plane; (B) Electrostatic field at a charged corner (all other charges assumed at "infinite" distance away); (C) Electrostatic field between oppositely charged slot and plane

(1) Draw the average potential line for region I. *
(2) Construct a normal to this line.
(3) Measure the angle between the path and the normal (the entrant angle).
(4) Calculate the average potentials for the two regions.
(5) Compute the new angle (β).
(6) Draw the segment of the trajectory for region II.
(7) Repeat for the other sections.

Segmented Parabolic Trajectories. A series of universal parabolas (Fig. 43), constructed by O. H. Schade, Sr., provide a method of constructing electron trajectories which is relatively easy to apply. This method is most generally used and probably gives the best results. Construction of the parabolas depends on the fact that the volt velocity of an electron is given by the potential value of its position,** regardless of the electron direction.

*An average potential line determines the average angle for a region. The average line is constructed, by eye, to fit midway between the known potential lines.

**Neglecting the velocity of emission and transit-time effects.

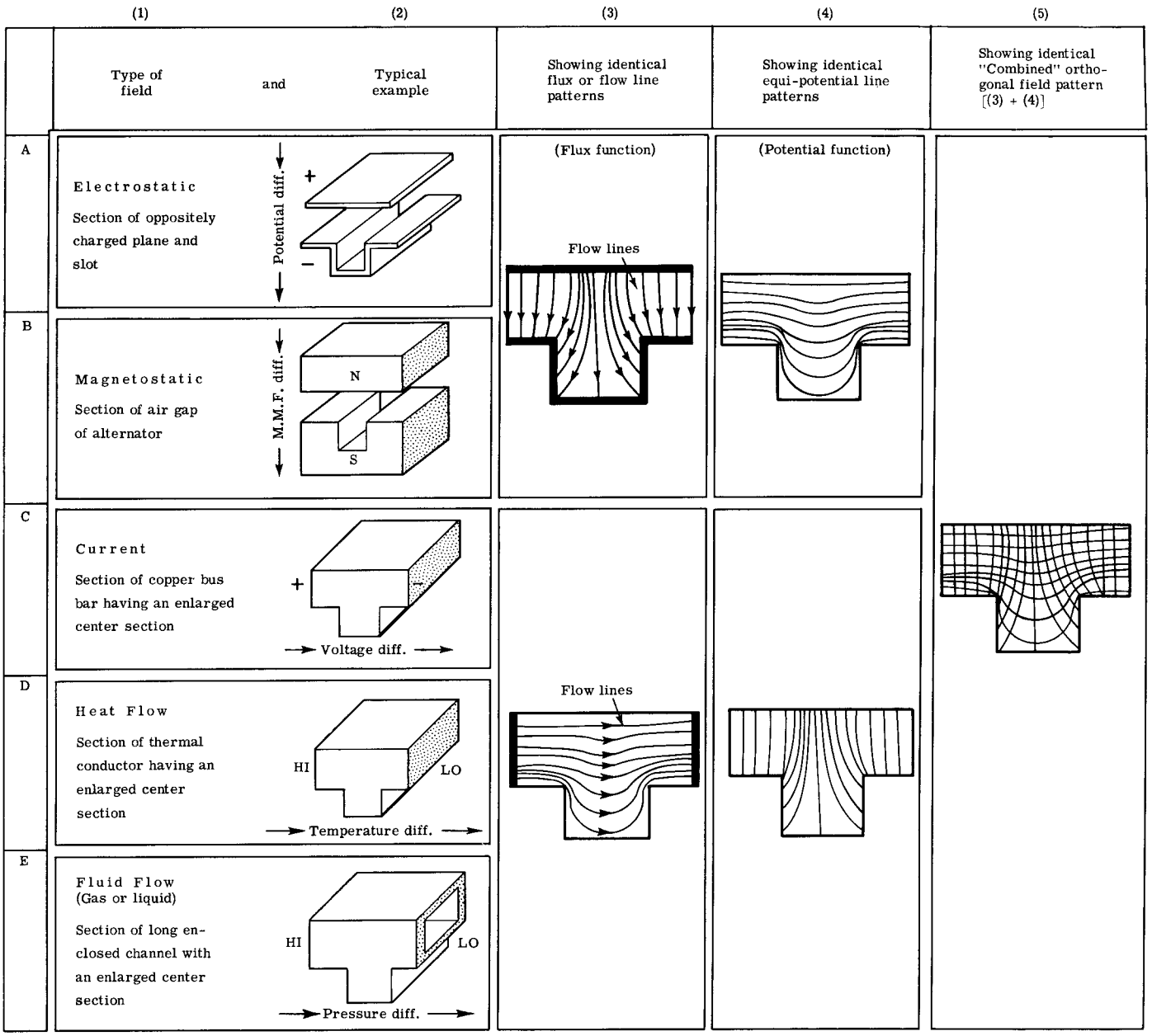


Figure 35. Analogies between Various Types of Fields

Construction is simplified by making the initial volt velocity of the electron V_0, when it leaves the potential line V_0, the unit in which the potential is measured (i.e., $V_0 = x_0$). If the electron leaves at an angle of 90 degrees with respect to the direction of the potential gradient, the electron path is given by the equation

$$y = 2\sqrt{x}$$

with x having the direction of the gradient in a right-angle coordinate system. (See Fig. 43 insert.) Both x and y are measured in units equal to V_0 (for the insert, V_0 is five divisions). The value of y_0 was made equal to $2V_0$ to simplify drawing. Assigning the unit V_0 various values results in the family of electron paths shown in Fig. 43.

Due to the fact that the electron volt velocity is equal to the value of the intersected potential line, the intersection of any electron path with a given potential line may be considered as the starting point of an electron with the velocity leaving at the angle shown. This type of plot results in a general family of curves showing the accurate path in a limen field for all electrons of all velocities leaving at all angles within the scale limits of the drawing.

The angle of incidence or slope of any curve is

$$\tan \alpha = \frac{dy}{dx} = \frac{1}{\sqrt{x}}$$

Thus, as the dimensions x_0, y_0, x, and y have the

same ratios for all points of equal slope due to the method of drawing, all points of equal slope are intersected by a straight line drawn through $x_o = y_o = 0$. Fig. 43 shows that any one of these lines is divided into equal sections by electron trajectories having equal increments of V_o, thus permitting correct interpolation of trajectories. The dotted line is a curved scale of points through which the slope indicating lines are to be drawn for the zero point. The 45-degree line is drawn as an example.

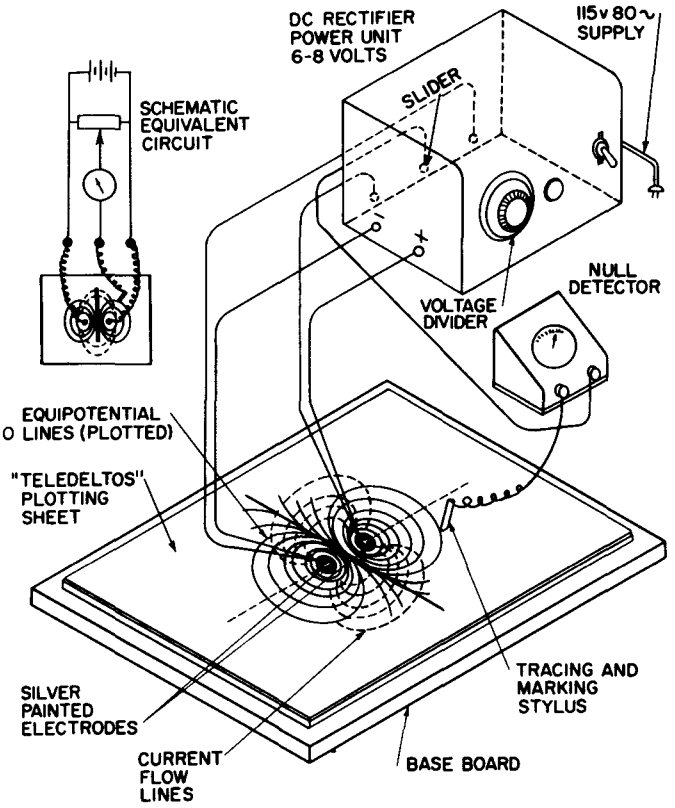


Figure 36. Components and Schematic Circuit of Analog Field Plotter

The use of the curves for tracing electron paths in field plots is illustrated in Fig. 44A. The path of an electron entering a linear field section at 30 volts with the angle shown is to be traced to the 80-volt potential line. The first step is to determine the location of the 0 potential line of this field. It is located at the distance $d_1 = (E_L/E_H - E_L)d_2$ if the linear field section is continued with the same gradient. The sketch is then placed over the family of curves of Fig. 43 with the 30-volt line at the distance d_1 from 0 and then shifted parallel to the potential lines until the angle of incidence of the electron matches one of the curves. The curve, which is then traced, represents the correct electron path.

If the equipotential lines are not straight and parallel (Fig. 44B) the parabolas may still be used; however, the field lines should first be adjusted. In many cases it is not advisable to subdivide the actual potential field into very short sections. For the case shown, the error of electron position is quite small if the average values of potential line direction and gradient between the limits A and B are used. Too fine a subdivision can cause cumulative matching errors unless it is done on a large scale and with great precision. Then again, if small differences of electron angles cause large computational deviations, the physical structure is very likely just as much affected by mechanical inaccuracies and requires high precision in manufacture to obtain optimum results.

Discussion

Trajectory tracing is a simple and mechanical, though tedious, procedure which can be handled most satisfactorily by digital computors if the volume of work warrants their use. Generally, in receiving-tube design, trajectories are a secondary consideration and are usually computed by hand. Special purpose analog computors have been built to plot trajectories directly in an electrolytic tank. The technique of rolling steel balls on a rubber sheet may also be used.[64, 65]

In any focused-beam device, the space-charge density may be quite large and greatly affect the trajectories. The solution is compensated, in these cases, by an integration procedure.

(a) Plot the field.
(b) Trace the beam.
(c) Recalculate the field from the knowledge of the approximate beam shape.
(d) Trace the new beam shape density, etc.
(e) Repeat as often as necessary to obtain the desired accuracy.

There is extensive literature on electron optics, beam devices, etc. This has originated in the fields where it is of primary concern and should be consulted.

Nomographs for the Rapid Modification of Electron-Tube Characteristics*

In the design and production of electron tubes, it is often necessary to modify internal tube dimensions to bring about desired changes in electrical characteristics. Three nomographs have been developed which make it possible to determine quickly the physical tube dimensions required to provide certain electrical characteristics, provided the dimensions and electrical characteristics of a tube of the same general class are already known. For instance, the tube design engineer can determine the internal spacings for a new amplifier tube which is to have higher transconductance than an existing type provided the dimensions and electrical characteristics of the existing type are known. In a similar manner, those engaged in electron-tube production could rapidly calculate the construction changes necessary to center various "off bogey" characteristics.

Theoretical Basis for Nomographs

It is assumed that the current in a diode is given by the simple Child-Langmuir law.

*The nomographs and the essential details of this section were supplied by Mr. R. D. Reichert of RCA Laboratories. See Ref. 66

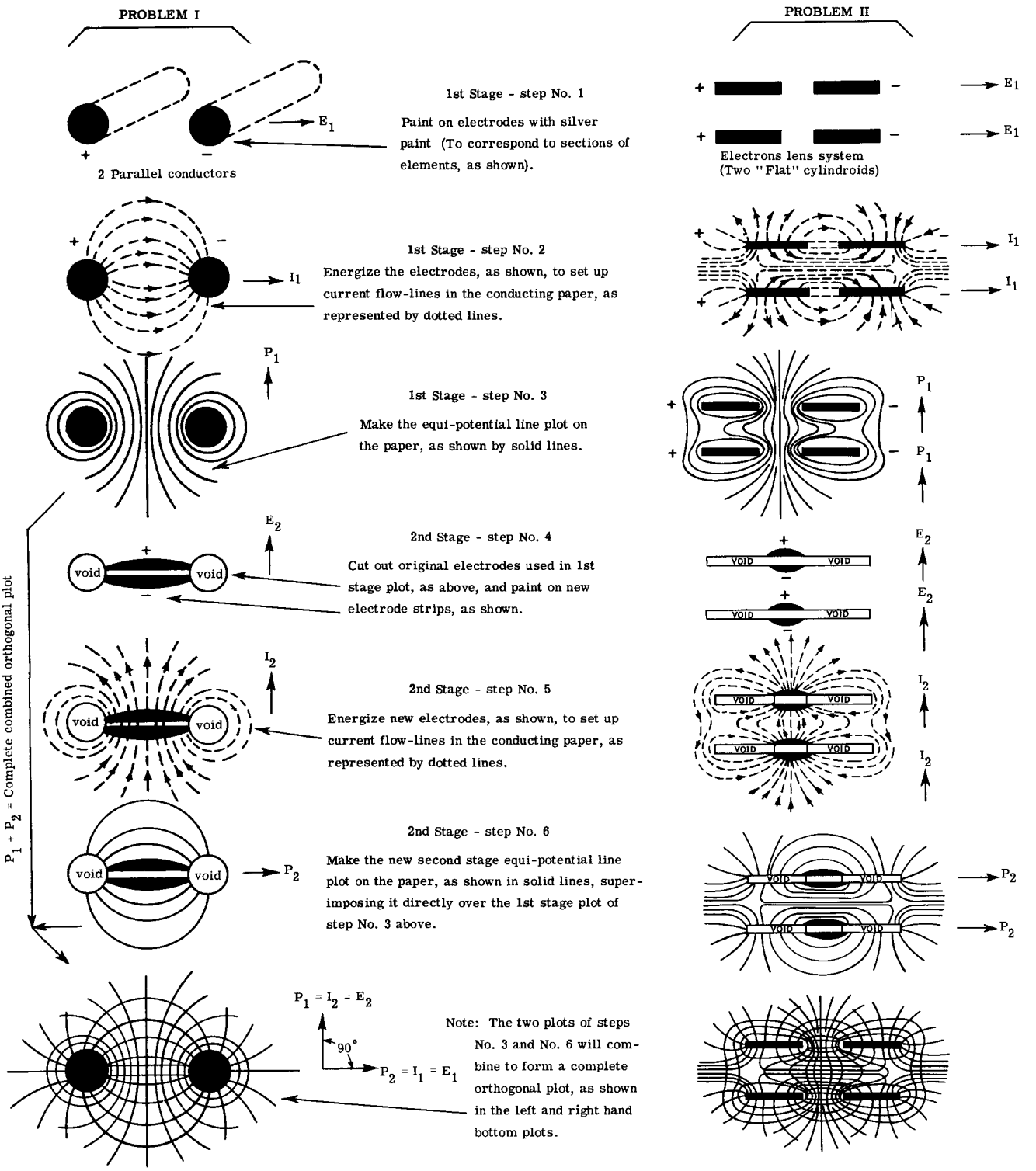

PROBLEM I
PROBLEM II
2 Parallel conductors
Electrons lens system
(Two "Flat" cylindroids)
1st Stage - step No. 1
Paint on electrodes with silver paint (To correspond to sections of elements, as shown).
1st Stage - step No. 2
Energize the electrodes, as shown, to set up current flow-lines in the conducting paper, as represented by dotted lines.
1st Stage - step No. 3
Make the equi-potential line plot on the paper, as shown by solid lines.
2nd Stage - step No. 4
Cut out original electrodes used in 1st stage plot, as above, and paint on new electrode strips, as shown.
2nd Stage - step No. 5
Energize new electrodes, as shown, to set up current flow-lines in the conducting paper, as represented by dotted lines.
2nd Stage - step No. 6
Make the new second stage equi-potential line plot on the paper, as shown in solid lines, superimposing it directly over the 1st stage plot of step No. 3 above.
P1 + P2 = Complete combined orthogonal plot
P1 = I2 = E2
P2 = I1 = E1
90°
Note: The two plots of steps No. 3 and No. 6 will combine to form a complete orthogonal plot, as shown in the left and right hand bottom plots.
void

Figure 37. Illustrations of "Two-Stage" Plotting Method

$$I = \frac{KA}{d^2} E_a^{3/2} \qquad (109)$$

This equation is extended to triodes by using a simple equivalent diode, as follows:

$$I = \frac{KA}{b^2} (E_a/\mu + E_c)^{3/2} \qquad (110)$$

and the gm is assumed to be given by the following expression:

$$gm = \frac{3}{2} \frac{KA}{b^2} (E_a/\mu + E_c)^{1/2} \qquad (111)$$

Empirical data has shown that the mu of a helical grid is given by the following law:

$$\mu = CN^2rs \qquad (112)$$

where C is a constant and N is the number of turns per inch of the lateral wire. For pentodes and tetrodes, if it is assumed that the anode voltage has a negligible effect, Eq. (110) may be rewritten using E_s and μ_{12}.

$$I_K = \frac{KA}{b^2}\left(\frac{Es}{\mu_{12}} + E_c\right)^{3/2} \qquad (113)$$

The term μ_{12} is often known as the triode mu T_μ. Because only a portion of the cathode current reaches the anode and, in the normal operating range, the ratio is constant, the current may be expressed as:

$$I_a = \frac{K'A}{b^2}\left(\frac{E_s}{T_\mu} + E_c\right)^{3/2} \qquad (114)$$

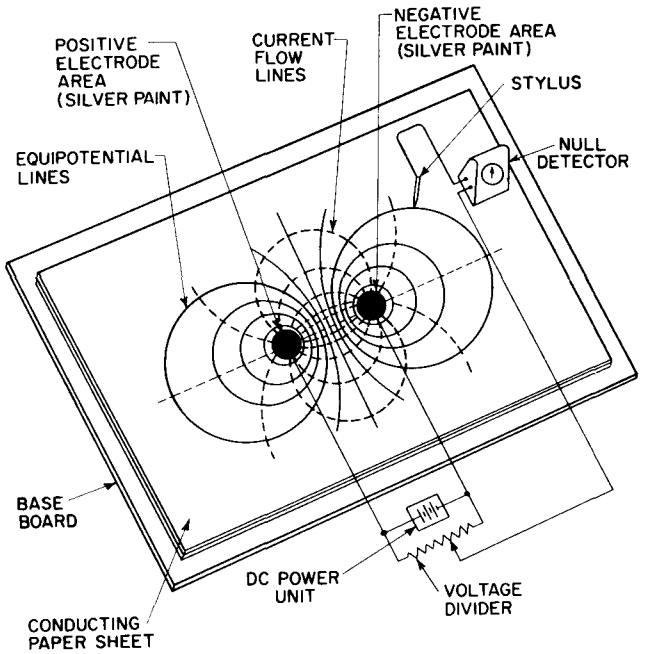


Figure 38. Typical Electrostatic Field Plot Showing Circuits, Electrodes, and Plotted Lines

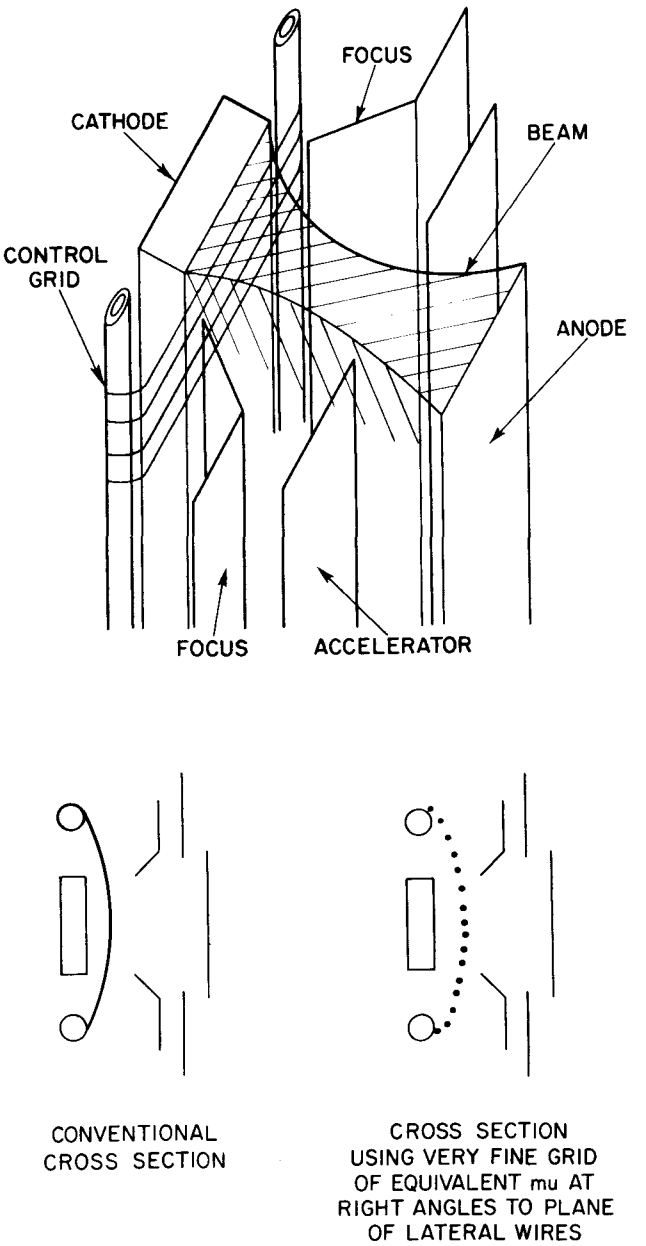


Figure 39. Method of Representing Grid for Field Plot

The previous equations do not take into account any of the phenomena attributable to close-spaced front ends (i.e., inselbildung effects). If Eq. (111) is divided by Eq. (110), the following results:

$$\frac{gm}{I} = \frac{3}{2}\,\frac{1}{E_a/\mu + E_c} \qquad (115)$$

This expression is not exactly true for real tubes, and the ratio varies if the ratio a/b is not held constant. Therefore, a factor is introduced which is dependent on a/b and the ratio gm/I is rewritten as follows:

$$\frac{gm}{I} = \frac{3}{2}\,F\,\frac{1}{E_a/\mu + E_c} \qquad (116)$$

Appplications and Limitations of the Nomographs

The nomographs (Fig. 45) described in this paper are used to determine the changes in tube dimensions that are required to modify the characteristics μ, gm, and I_a. The starting point in the use of the nomographs is always the notation of data from an existing comparable type (referred to as the original tube) on the appropriate axis of the nomograph. This operation determines various tube factors and constants. The next step is the notation of the desired electrical characteristics while the factors determined by the original tube are kept constant. The intersection points on certain axes then indicate the dimensions of the desired tube. The nomographs cannot be used for changing the ratio of plate current to screen current of pentodes, modifying the suppression characteristics of pentodes, or for modifying interelectrode capacitances.

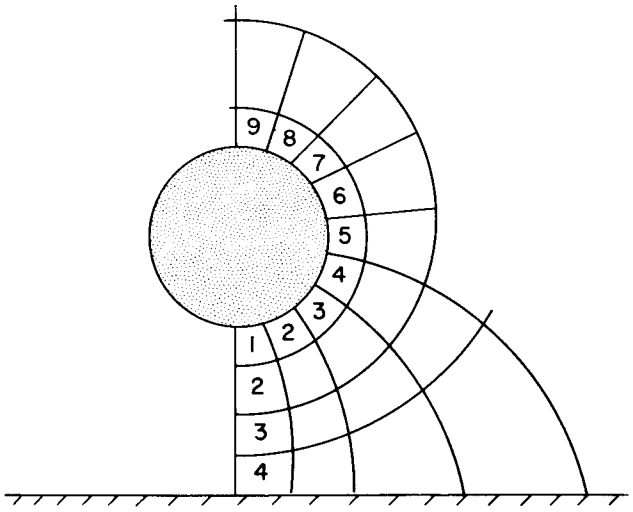


Figure 40. Field Sketching Using Curvilinear Squares

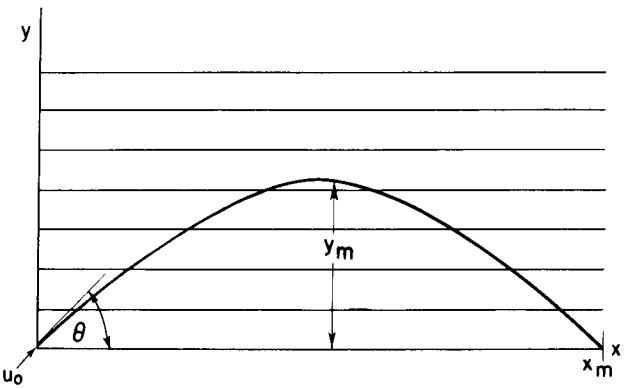


Figure 41. Electron Injection into a Uniform Field

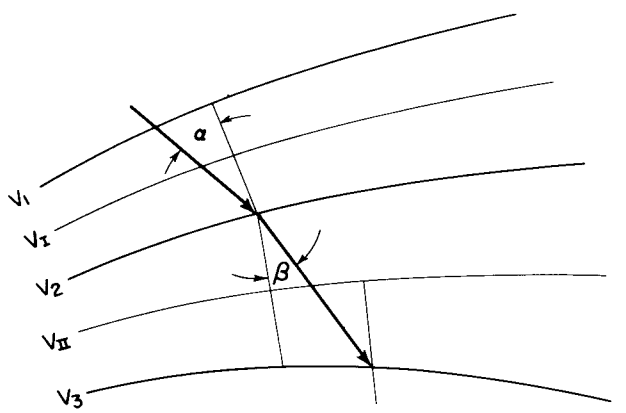


Figure 42. Example of the Snell's Law Method of Trajectory Tracing

The use of the nomographs assumes that the factor F and the constants K and C remain constant (or vary so little that the variations may be neglected) over the range of the original tube and the desired tube.

Many of the modifications of tube characteristics may be accomplished in more than one way. For instance, diode plate current at a given plate voltage may be increased by either decreasing plate-cathode spacing or by increasing the cathode area. In all cases of multiple solutions, the nomographs provide either solution and, in addition, many solutions which represent various combinations of the two. The user of the nomographs must decide on the desired solution by considering other requirements such as interelectrode capacitances, available heater input power, and the ease with which either change can be made (an important factor when the tube being modified is in production).

The nomographs (and the equations upon which they are based) are valid only over a limited range. Although extreme solutions may be obtained which appear mathematically correct, it is impractical or impossible to construct a useful vacuum tube based on these extreme solutions. In general, if the grid-cathode spacing factor a/b and the screening factor 2r/a of the desired tube are kept similar to the corresponding factors of the original tube, the solution obtained is valid.

In triodes and pentodes, the designer does not have complete freedom in the choice of I_a, gm, and triode μ. Eq. (115) shows that for a given set of applied voltages, two of the three above characteristics may be chosen but the third characteristic is then dependent. When solving triode and pentode problems, by using the nomographs, no attempt should be made to modify I_a, gm, and triode mu simultaneously. Instead, two of these characteristics should be modified and the problem solved with the nomographs to see where the third characteristic falls. If the third characteristic is a value which is other than that desired, the problem becomes one of an advanced nature and requires a significant modification of the grid-cathode spacing factor a/b which, in turn, will modify the F factor.

In the nomographs, straight lines are drawn intersecting three axes at a time. The axes marked "bend" indicate turning points. In the case of diode problems, it is only necessary to use axes V through XI of nomograph (B). It is not always possible to work across from left to right. For some problems, the user has to start at both ends independently; however, a continuous line bent at the "bend" axes should be the final result.

The basic steps for using the nomographs are as follows:

(1) Enter into the nomographs the known characteristics and dimensions of a tube which is in the same general class as the desired tube.
(2) Keep F and K constant (i.e., passing through the same intersection points on the F and K axes); enter the desired electrical characteristics of the new tube.
(3) Read the dimensions of the new tube; obtain practical solutions when more than one solution is possible.

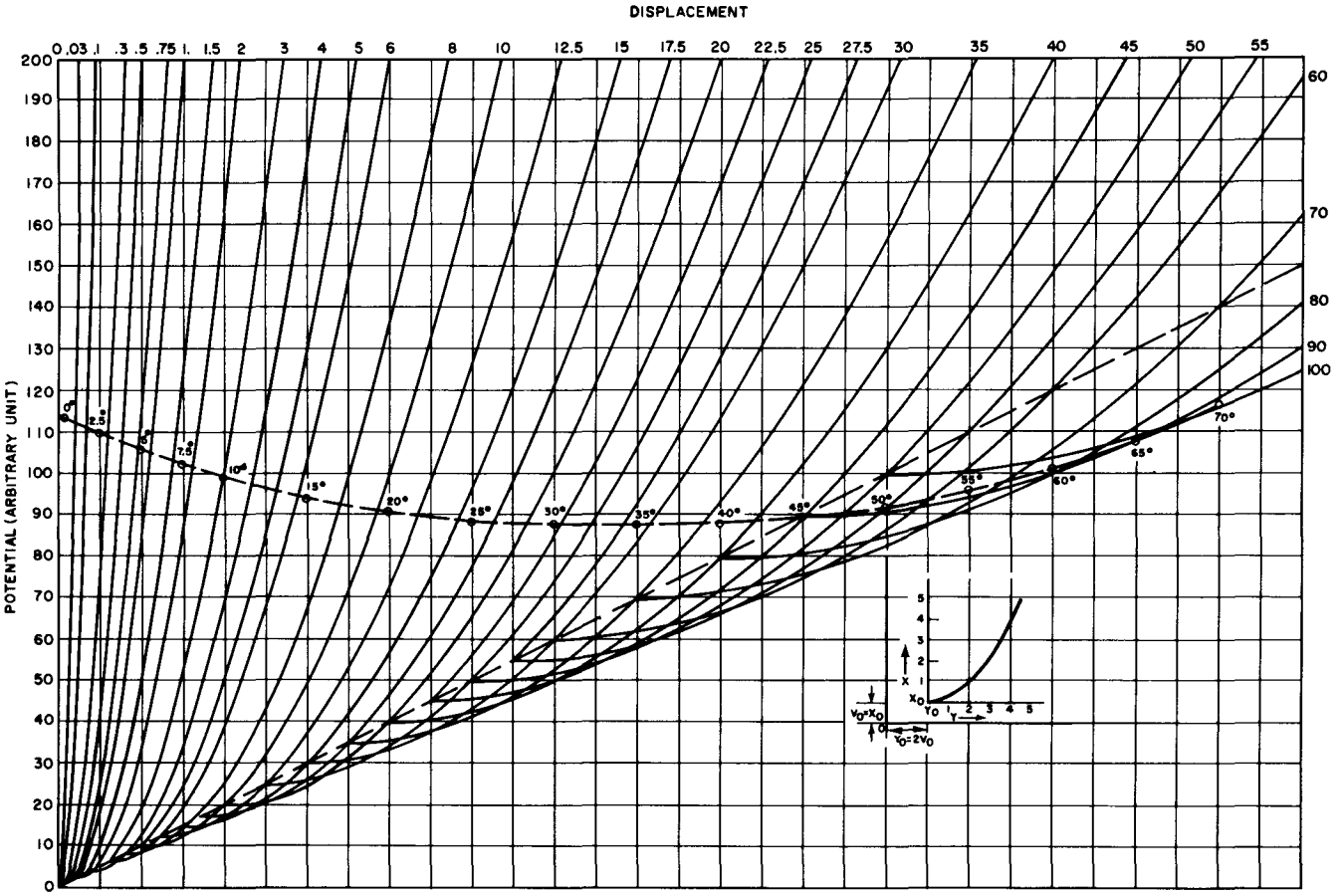


Figure 43. O. H. Schade's Parabolas

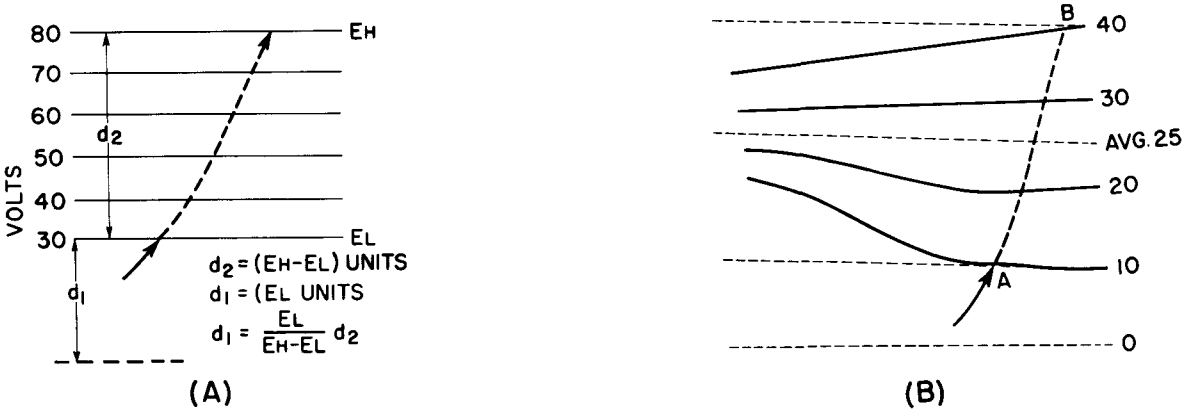


Figure 44. Example of the Piece-Wise Parabolic Method of Trajectory Tracing: (A) Plot of electron path in linear field; (B) Plot of electron path in nonlinear field

Use of Multiplying Factors on Nomographs

Any of the scales on nomographs (B) and (C) may be multiplied by any factor provided the same factor is used throughout the entire problem. The multiplication of any axis, or combinations of axes, by any factors does not require the correction of other axes.

In the case of nomograph (A), however, it is not possible to apply multiplying factors to the various scales. If a particular problem involves values which are beyond the scales of nomograph (A), the value of $(E_a/M) + E_c$ can be calculated arithmetically and inserted in nomograph (B).

REFERENCES

1. Kellog, O. D., Foundations of Potential Theory, Julius Springer, Berlin, 1929

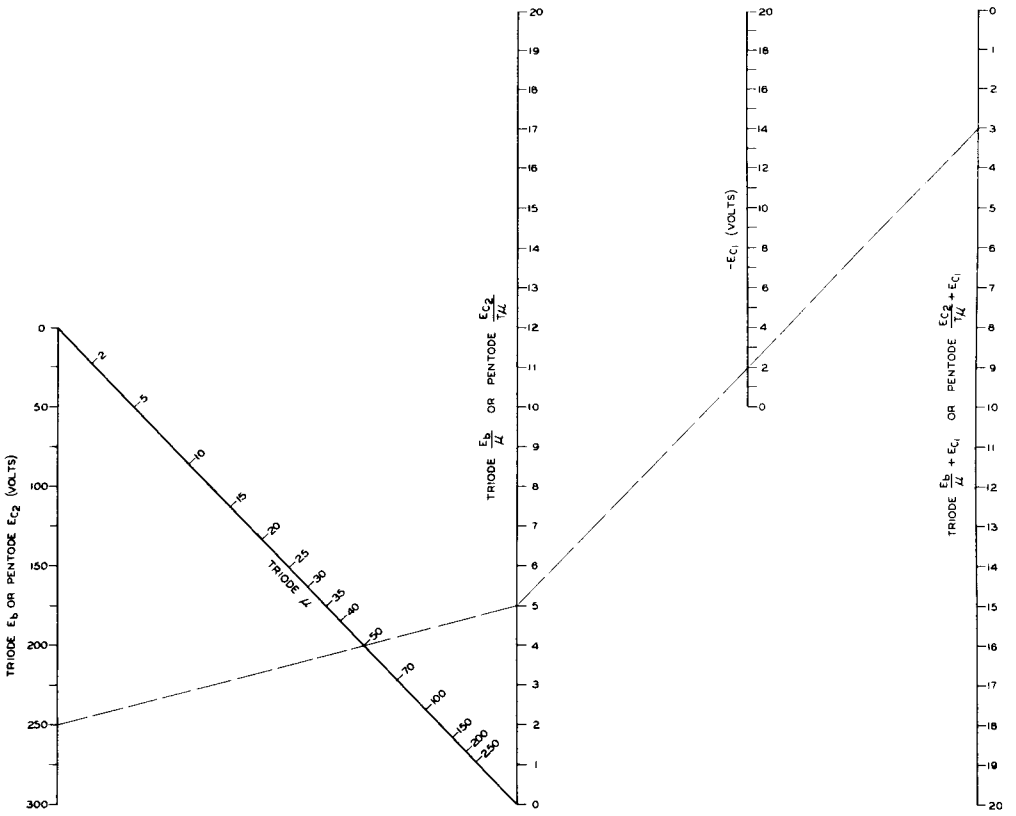

TRIODE E_b OR PENTODE E_{c_2} (VOLTS)
TRIODE μ
TRIODE $\frac{E_b}{\mu}$ OR PENTODE $\frac{E_{c_2}}{\mu}$
$-E_{c_1}$ (VOLTS)
TRIODE $\frac{E_b}{\mu} + E_{c_1}$ OR PENTODE $\frac{E_{c_2}}{\mu} + E_{c_1}$

Figure 45A. Nomograph for the Rapid Adjustment of Tube Characteristics

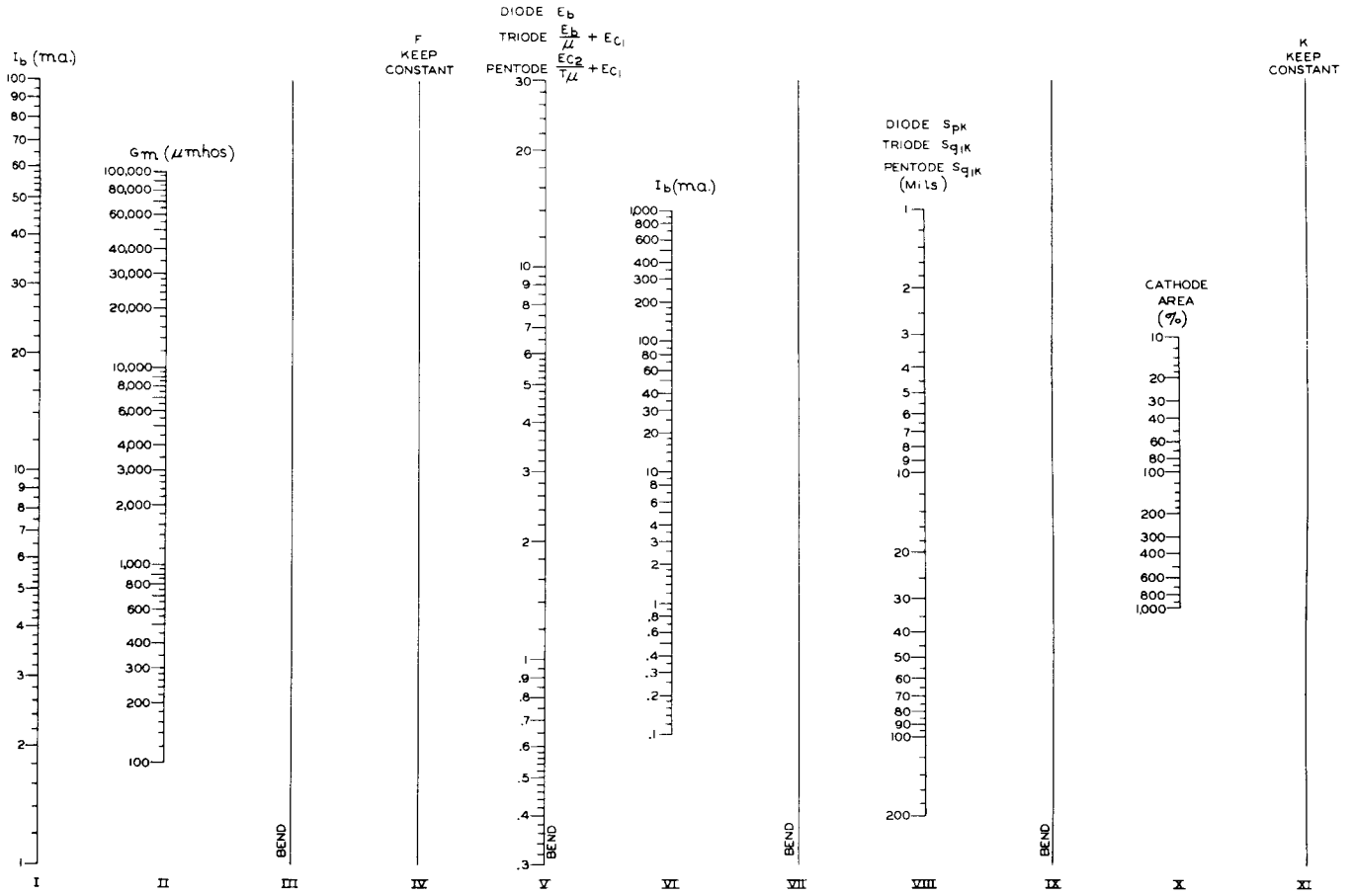


Figure 45B. Nomograph for the Rapid Adjustment of Tube Characteristics

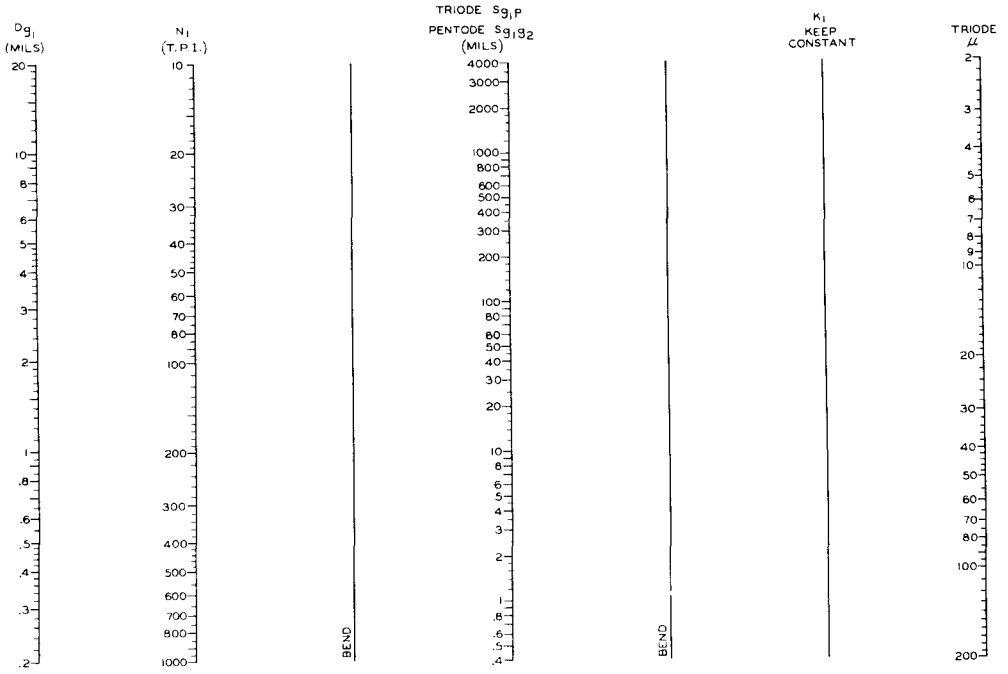


Figure 45C. Nomograph for the Rapid Adjustment of Tube Characteristics

2. Ramo, S., and J. R. Whinnery, Fields and Waves in Modern Radio, John Wiley and Sons, 2nd Edition, 1953
3. Stratton, J.A., Electromagnetic Theory, McGraw-Hill, 1941
4. Salzberg, B., and A. V. Haeff, "Effects of Space Charge in the Grid-Anode Region of Vacuum Tubes," RCA Review, Vol. 2, pp. 336-374, 1938
5. Fay, C.E., A. L. Samuel, and W. Shockley, "On the Theory of Space Charge Between Parallel Plane Electrodes," Bell Syst. Tech. Jour., Vol. 17, pp. 49-79, January 1938
6. Ramberg, E. G., and L. Malter, "Potential Distribution for Space-Charge-Limited Current Between a Plane Accelerating Grid and Parallel Anode," Jour. Appl. Phys., Vol. 23, pp. 1333-1335, December 1952
7. Beck, A. H. W., Thermionic Valves, Cambridge Univ. Press, 1953
8. Langmuir, I., "The Effects of Space Charge and Residual Gases on Thermionic Currents in High Vacuum," Phys. Rev., Vol. 2, pp. 450-486, 1913
9. Childs, C.D., Phys. Rev., Vol. 32, p. 492, 1911
10. Ivey, H. F., "Space Charge Limited Currents," Advances in Electronics and Electron Physics, Vol. VI, pp. 137-256 (contains 215 refs.), 1954
11. Millman, J., and S. Seely, Electronics, McGraw-Hill, 2nd Edition, 1951
12. Langmuir, I., "The Effect of Space Charge and Initial Velocities on the Potential Distribution and Thermionic Current Between Parallel Plane Electrodes," Phys. Rev., Vol. 21, p. 419, 1923
13. Langmuir, I., and K. T. Compton, "Electrical Discharges in Gases," Part II, Rev. of Mod. Phys., Vol. 3, p. 191, 1931
14. Kleynen (Kleijnen), P. H. J. A., "Extension of Langmuir's (ξ , η) Tables for a Plane Diode with a Maxwellian Distribution of the Electrons," Philips Res. Rpt., Vol. 1, pp. 81-96, 1945
15. Lieberman, G., "The Calculation of Amplifier Valve Characteristics," Jour. IEE, Part III, Vol. 93, pp. 138-153, 1946
16. Ferris, W. R., "Some Characteristics of Diodes with Oxide Coated Cathodes," RCA Review, Vol. 10, p. 134, 1949
17. Langmuir, I., and K. B. Blodgett, "Currents Limited by Space Charge Between Coaxial Cylinders," Phys. Rev., Vol. 22, p. 347, 1923
18. Maxwell, J. C., A Treatise on Electricity and Magnetism, 2 Vols., Oxford, 3rd Edition, 1904
19. Kleijnen, P. H. J. A., "The Penetration Factor and the Potential Field of a Planar Triode," Philips Res. Rpt., Vol. 6, pp. 15-33, February 1951
20. Hsu, H., and C. E. Horton, "Electrolytic Tank Measurements of Mesh Grid Characteristics," IRE Nat. Convention Rec., Part 3, Vol. 5, pp. 114-119, 1957
21. Vodges, F. B., and F. R. Elder, "Formulas for Amplification Factors for Three Element Tubes," Phys. Rev., Vol. 24, pp. 283-289, Dec. 1924
22. Herne, H., "Valve Amplification Factors," Wireless Eng., Vol. 21, pp. 59-64, February 1944
23. Ollendorff, F., "Calculation of Controllance Due to a Fine-Mesh Grid," Electrotechnik und Maschinenbau, Vol. 52, pp. 585-591, 1934
24. Thompson, B. J., "Space Current Flow in Vacuum Tube Structures," Proc. IRE, Vol. 31, pp. 485-491, September 1943
25. Walker, G. B., "Theory of Equivalent Diode," Wireless Eng., Vol. 24, pp. 5-7, 1947
26. Telegren, B. D. H., "De Grootte van de Emissistroom in enn Triode," Physica, Vol. 5, pp. 301-315, 1925

27. Llewelyn, F.B., and L. C. Peterson, "Vacuum-Tube Networks," Proc. IRE, Vol. 32, pp. 144-166, 1944
28. Strutt, M. J. O., Elektronenröhren, Springer-Verlag, Berlin, 1957
29. Schade, O. H., "Beam Power Tubes," Proc. IRE, Vol. 26, pp. 137-180, 1938
30. Rothe, H., and W. Kleen, Hochvacuum-Electronen-Rohren, Band 1, Physikalische grundlagen, Akademische Verlagsgesellschaft M. B. H., Frankfurt a.m., 1955
31. Jonker, J. L. A., and B. D. H. Tellegren, "The Current to a Positive Grid in Electron Tubes," Philips Res. Rpt., Vol. 1, pp. 13-32, 1945
32. Jonker, J.L.A., "The Control of the Current Distribution in Electron Tubes," Philips Res. Rpt., Vol. 1, pp. 331-338, November 1946
33. Deb, S., "Current Distribution in Plane Positive-Grid Triodes," Indian Jour. Phys., Vol. 26, pp. 337-392, August 1952
34. Spangenberg, K. R., Vacuum Tubes, 1st Edition, McGraw-Hill, 1948
35. Spangenberg, K., "Current Division in Plane Electrode Triodes," Proc. IRE, Vol. 28, pp. 226-236, May 1940
36. Jonker, J.L.H., and Z. van Gelder, "The Internal Resistance of a Radio-Frequency Pentode," Philips Res. Rpt., Vol. 12, pp. 141-175, April 1957
37. Bennett, W.R., and L. C. Peterson, "The Electrostatic Field in Vacuum Tubes with Arbitrarily Spaced Elements," Bell Syst. Tech. Jour., Vol. 28, pp. 303-314, 1949
38. Dahlke, W., "Gittereffectivpotential und Kathodenstromdichte einer Ebenen Triode unter Berücksichtigung der Inselbildung," Telefunken Zeit., Vol. 24, pp. 213-222, December 1951
39. Napiorski, F. R., O. H. Schade, Jr., and L. C. Scholz, Unpublished work
40. Harris, W.A., and H.W. Kirk, Unpublished work
41. Dahlke, W. A., "Statische Kennwerte der Nichtidealen Triode," Telefunken Zeit., Vol. 27, pp. 172-187, September 1954
42. Weber, E., "Mapping of Fields," Electromagnetic Fields, Vol. I, John Wiley and Sons, 1950
43. Kober, H., Dictionary of Conformal Representations, Dover Publications, 1952
44. Breckenback, E.F. (ed), "Constructions and Applications of Conformal Maps," Nat. Bur. of Standards, Appl. Math Series, Vol. 18, 1952
45. Ince, E.L., Ordinary Differential Equations, 4th Edition, Dover Publications
46. Bateman, H., Partial Differential Equations, Cambridge University Press, 1932
47. Southwell, R. V., Relaxation Methods in Engineering Science, Oxford University Press, 1940
48. Southwell, R.V., Relaxation Methods in Theoretical Physics, 2 Vols., Oxford University Press, 1946, 1956
49. Liebmann. G., "Electrical Analogies," Brit. Jour. Appl. Phys., Vol. 4, p. 193, 1953
50. Harries, J. H. O., "The Rubber Membrane and Resistance Paper Analogies," Proc. IRE, Vol. 44, pp. 236-248, February 1956
51. Kennedy, P. A., and G. Kent, "Electrolytic Tank, Design and Applications," Rev. of Sci. Inst., Vol. 27, p. 916, November 1956
52. Kleynen, P. H. J. A., "The Motion of an Electron in Two-Dimensional Electrostatic Fields," Philips Tech. Rev. Vol. 2, pp. 338-345, 1937
53. Code W1
54. Code W2
55. Francken, J. C., "The Resistance Network, a Simple and Accurate Aid to the Solution of Potential Problems," Philips Tech. Rev., Vol. 21, pp. 10-23, 1959-1960
56. Alma, G.A., G. Diemser, and H. Groendijk, "A Rubber Membrane Model for Tracing Electron Paths in Space Charge Fields," Philips Tech. Rev., Vol. 14, pp. 336-343, 1953
57. Sunshine Scientific Co., "Use and Theory of the Analog Field Plotter," Instruction Manual for Apparatus Manufactured by the Sunshine Co.
58. Einstein, P. A., "Factors Limiting the Accuracy of the Electrolytic Plotting Tank," Brit. Jour. Appl. Phys., Vol. 2, p. 49, 1951
59. Van Duzer, T., and G. R. Brewer, "Space-Charge Simulation in an Electrolytic Tank," Jour. Appl. Phys., Vol. 30, pp. 291-301, March 1959
60. Code W3
61. Stanford Electronics Laboratories, "Electron Devices Research," Consolidated Quarterly Report No. 10, April 1-June 30, 1959, pp. 45-47, Project 314T, Electrolytic Tank Studies, Contract No. DO36(039)SC73178
62. Pierce, J. R., Theory and Design of Electron Beams, D. Van Nostrand and Co., Inc., 1954
63. Harman, W. W., Fundamentals of Electron Motion, McGraw-Hill, 1953
64. Klemperer, O., Electron Optics, Cambridge University Press, 1939
65. Liebmann, G., "Field Plotting and Ray Tracing in Electron Optics. A Review of Numerical Methods," Advances in Electronics, Vol. II, pp. 101-149, Academic Press, 1950
66. Code W4
67. Dingwall, A. G. F., "The Mathematical Basis for the Computer Program," Appendix B, Quarterly Reports 1 and 2, "A Study of Electron Tube Deterioration Utilizing Kinetic Theory," Contract No. NObsr-77637. (This is a discussion of the generalized Child-Langmuir equation referred to above. The derivation of the correction factor $(\theta\eta)$ and of the general form of the equations was done by A. G. F. Dingwall, Jr., but has not yet been published.)
68. Fremlin, J., "Calculation of Triode Constants," Elec. Communications, Vol. 18, pp. 33-49, July 1939
69. Fremlin, J. H., R. N. Hall, and P. A. Shatford, "Triode Amplification Factors," Elec. Communications, Vol. 23, pp. 426-435, Dec. 1946
70. Rothe, R. (ed), F. Ollendorff, and K. Pohlhausen (trans. A. Herzenberg), Theory of Functions, As Applied to Engineering Problems, Technology Press, MIT, 1933
71. Richardson, O. W., The Emission of Electricity from Hot Bodies, Longmans Green, 2nd Edition, 1921

Space Current Flow

M. V. Hoover

Lancaster

THE ELECTRON

The electron[1] apparently embodies the fundamental electrical charge of the universe. Proof of its existence and the measurement of its physical attributes of magnitude by experimental means constitute the keystone of modern science. As early as 1873, Maxwell referred to the electron as the "molecule of electricity." The electron is said to be the "planetary unit" of all atoms. Matter is dependent upon electrons for its existence, but electrons apparently can exist by themselves. "Free electrons" are responsible for most electrical phenomena. They are the vehicles which constitute the flow of traffic in vacuum tubes. Their motion in conductors is measured in terms of a "current." Electron migrations also occur in semiconductor devices like transistors. Their oscillatory movements in special conductors such as antennas give rise to electromagnetic radiations known as radio waves. The atomic physicist testifies that his "beta particles" are high-speed electrons.

Visualization of the electron is hampered by its enigmatic nature. Modern engineering depicts the electron as a particle of matter having low mass when at rest or at low velocities. Physicists, on the other hand, find this picture of the electron far from adequate since there are some applications in which the electron displays more of a wave aspect than a particle aspect. This is the case, for example, in the electron microscope, where a high-velocity beam of electrons acts as though it were a "ray" of very short wave length, somewhat akin to light "rays." The confusion is further augmented by the discovery of the spinning electron and the suggestion by DeBroglie that the electron is some sort of complex electromagnetic wave motion. A school of wave mechanics holds that the electron is not simply a charged mechanical mass point. Their representation is presumably best described by a complicated mathematical expression which accurately predicts electron behavior but for which there is no simple mechanical picture.

In most applications of practical electronics, the electron may be visualized as the lightest "particle" known despite the fact that it has an apparent density of 0.5×10^{11} grams per cubic centimeter which is more than a billion times greater than that of the heaviest metals (the density of uranium is about 19 grams per cubic centimeter). In practice, electrons may also be visualized as the "carriers" of electric charge, the measured quantity of electric charge carried by each electron being

$$q = 1.591 \times 10^{-19} \text{ coulomb}$$

$$e = 4.77 \times 10^{-10} \text{ electrostatic units}$$

Thus, for a current of one ampere flowing through a wire (one ampere equals one coulomb per second), the number of electrons per second which pass a given cross section of the wire is given by

$$n = \frac{1}{1.591 \times 10^{-19}} = 0.629 \times 10^{19}$$

An electron is also the origin of a field of force and, therefore, has an equivalent mass which, expressed in grams for the electron at rest, is

$$m = 9.035 \times 10^{-28} \text{ grams}$$

If it is assumed that the electron has spherical shape, the size of its apparent radius is given by

$$r = 1.85 \times 10^{-13} \text{ centimeter}$$

If an electron is moved through a potential difference of V' electrostatic units (e.s.u.), its kinetic energy changes according to the relation

$$V'q = 1/2\ mv^2$$

Changing from e.s.u. to volts (one e.s.u. equals 300 volts) yields the following expression which describes the velocity in terms of voltage

$$v = 5.93 \times 10^5 \text{ volts (v is in meters per second)}$$

EMISSION OF ELECTRONS[2]

The functioning of every electron tube is dependent upon the generation and control of free electrons which act as carriers of charge in establishing a current. Equally necessary is the cathode electrode from which the electrons emanate as a consequence of the mechanism of emission. It is possible to increase the energy of the electrons in a cathode to a point where they become sufficiently energetic to leave the parent cathode and escape into the free space adjacent to the cathode. They are then said to have been emitted. The five most

common methods of producing such emission are:

1. Thermionic, or primary, emission
2. Secondary emission
3. Photoelectric emission
4. High-field emission
5. Radioactive disintegration

The types of emission differ only in the mechanism by which the escape energy is imparted to the free electrons in a cathode. Once electrons have been liberated from a cathode electrode they obey similar laws regardless of the mechanism by which they were emitted. With the exception of electrons produced by radioactive disintegration, electrons are most easily emitted in vacuum. All types of emitted electrons are also most effectively utilized in vacuum. Electrons emitted in air are impeded by the relatively dense surrounding atmosphere.

Thermionic, or Primary, Emission

Thermionic emission is analogous to the evaporation of a liquid. Just as heat energy is required to evaporate a liquid, so is heat energy required to "evaporate" thermionic electrons. Davisson and Germer were able to show that the thermionic emission of electrons even causes cooling of the cathode in a manner analogous to the cooling associated with the evaporation of a liquid. At normal temperatures, the molecules of a liquid have some thermal motion but only a few molecules are sufficiently energetic to "jump out" far enough to leave the surface. Their motion is more violent at increased temperature; therefore, an increasing number do overcome the attraction of the liquid and do evaporate or "boil out." In a similar manner, the electrons in a cathode are closely held at ordinary temperatures by a type of "electron affinity." An increase in temperature imparts increasingly greater energy to an increasingly greater number of electrons so that they may have sufficient velocity to escape from the cathode surface. As the electron (negative charge) leaves the cathode surface, it leaves its image (positive charge) which attracts the electron back toward the surface. As the newly freed electron moves into space, the force of attraction from the positive image charge in the cathode tends to balance the initial electron escape velocity and restrain its movement sufficiently so that the electron cannot stray very far from the surface of the heated solid.

Assume a cathode to be heated in a vacuum so that oxidation cannot occur. At a certain temperature of the cathode, the motion of the molecules and electrons in the cathode becomes great enough so that a number of electrons are emitted. A further rise in cathode temperature will be accompanied by an increase in the kinetic energy of the electrons and an increase in the number and initial velocity of those emitted. Finally, a thick atmospheric cloud of electrons adjacent to the hot cathode is produced. In the absence of external electric fields, a space-charge cloud of electrons is thus formed.

It is possible to determine experimentally the energy that an electron must attain to be emitted from any material. This electron affinity is usually alluded to as the work function of the cathode. Exacting physical studies of the work function of cathodic materials have been notably fruitful in the quest for fundamental physical knowledge. Tungsten, for example, has been found to have a thermionic work function of 4.52 volts; i.e., the electron must possess a kinetic energy in the amount of 4.52 joules per coulomb to escape from its tungsten cathode; or, since an electron has a charge of 1.602×10^{-19} coulomb, 7.241×10^{-19} joule is required for the thermionic emission of an electron from tungsten. Other materials have different work function values, some higher, some lower. One might at first assume that the material having the lowest work function would be the most suitable as a cathode, but this assumption is not necessarily true because, in addition, the cathode must possess certain mechanical characteristics. Two outstanding cathode materials having lower work functions than tungsten are thorium (3.4 volts) and calcium (3.2 volts), but both vaporize at temperatures that are too low to produce copious emission. The choice of a suitable cathode material involves the consideration of many factors, among which are a low work function, a high melting point, and a long life.

No completely successful theoretical derivation of the emission phenomenon has apparently as yet been made. Richardson and Dushman have summarized the situation by means of the following expression which approximately describes the effects of the various variables for the case of thermionic emission from a metal:

$$J = AT^2 \epsilon^{-b_0/T} \qquad (1)$$

where J = current density, amperes per cm^2
A = 120.4 amperes per cm^2 per deg^2, a universal theoretical constant
T = absolute temperature in degrees Kelvin (equal to 273+ degrees Centigrade)
b_0 = temperature equivalent of the work function, $11{,}600\,\phi_0$ in degrees Kelvin, where ϕ_0 is the work function of the metal in volts
ϵ = the base of natural logarithms, 2.71828

The exponential term in the emission equation depicts the variation of emission with temperature. Variation with the T^2 term is so small that the accuracy of the exponent (2) can hardly be verified experimentally. Thus, in the case of tungsten operating at 2500 K, a one per cent change in temperature changes the T^2 term by only two per cent but changes the exponential term by twenty per cent. This expression illustrates the fact that the emission-temperature function is one of the most rapidly varying functions found in nature. Doubling the temperature may increase the emission by a factor of 10^7.

Fortunately, it is possible to raise some metals to temperatures higher than their melting temperatures in the pure state by using them in various chemical and physical combinations. Thus, a monatomic layer of thorium on tungsten can be operated at or above the melting temperature of thorium itself. Furthermore, it has been found that small bits of the pure metal can be made to diffuse out of an oxide in the case of the rare-earth metals so that advantage can be taken of the low work function of these metals, which would otherwise

melt at low temperatures. From the above remarks, it can be seen that three classes of emitters exist: (1) pure metals (e.g., tungsten), (2) atomic-film emitters (e.g., thoriated tungsten), and (3) oxide emitters (alkaline-earth oxide coatings like BaO, SrO, and CaO).

Many scientists and engineers have devoted their lives to the study of these three classes of emitters. Interested readers are referred to the numerous volumes in the literature which record the findings and theories of these scientific laborers.

Secondary Emission

Secondary emission is the process whereby electrons are ejected from a solid or liquid as a consequence of bombardment by electrons, positive ions, or other rapidly moving particles. Secondary emission usually is induced by the impingement of primary electrons on an electrode of positive potential. Under these circumstances a single primary impinging electron may eject from one to ten secondary electrons, depending upon the work function of the material, the condition of its surface, the angle of incidence, and the accelerating voltage (which determines the velocity of the impinging electron). It is presumed that the kinetic energy of the primary electron is imparted to the secondary electrons and added to their normal thermal kinetic energy in such a way that they are able to overcome the potential barrier or work function of the surface. In essence, the energy that enables the free electrons to escape comes from the striking particle.

Secondary emission does not constitute a violation of the law of conservation of energy, since the sums of the energies of the secondary electrons are always lower than those of the primary electrons. Secondary electrons apparently are not produced for primary velocities below about nine volts. As the potential of the bombarding primary electrons is increased, the ratio of secondary-to-primary electrons increases, reaching a maximum in the vicinity of about 400 volts for all metals, and then decreases with increasing bombardment voltage. Most of the metals reach a maximum ratio of secondary electrons to primary electrons of between one and three; some of the complex alkali metals reach ratios of as high as ten. Maximum ratios depend greatly upon processing and surface conditions. The number of liberated secondary electrons depends greatly upon the potential conditions surrounding the bombarded surface. If there is an adjacent electrode at a potential higher than that of the bombarded surface, it will collect most of the secondary electrons. If the adjacent electrode has a potential lower than that of the bombarded surface, only the highest velocity secondaries will go to adjacent electrodes, and the low-velocity electrons will return to the bombarded surface from which they came.

Secondary emission is usually regarded as a nuisance in most electron tubes with the exception of devices like electron multipliers, magnetrons, and various types of camera tubes which are all dependent upon the phenomenon for their operation. Secondary emission exists at the anodes of all vacuum tubes. Usually, the secondaries thus emitted are attracted back into the anode by its positive potential and therefore have little effect upon the operation of the tube; but if there happens to be a second electrode, juxtaposed to the source of secondary emission, which has a higher positive potential than this source, the secondaries will tend to flow to the second electrode. This effect may readily happen in the four-electrode (tetrode) tube, at times to the detriment of its performance. In the case of the screen-grid electrode in tetrodes it is not unusual to find that the flow of secondary electrons leaving the screen grid is actually greater (under certain conditions) than the flow of primary electrons to the screen grid. These effects of secondary emission in vacuum tubes may be obviated by the insertion of an additional, negative electrode between the two positive ones. The negative potential on this electrode will drive secondary electrons back into the surface from which they were emitted and thus nullify the harmful effect. Note, however, that the secondary emission is not eliminated, the negative electrode merely prevents the secondaries from migrating to other electrons where they might interfere with the operation of the tube. Pentode vacuum tubes illustrate the practical use of this technique. Secondary emission also occurs in cathode-ray picture tubes when the beam electrons strike the fluorescent screen; the collection of these secondaries is a means of completing the circuit for the flow of current. In essence, secondary emission is commonly encountered in multiple-electrode tubes, where it has the effect of somewhat altering the normal primary electron current characteristics.

Although the complete theory of secondary emission is not as yet completely understood, there is a vast accumulation of observed effects. In summary, the number of secondary electrons emanated depends on: (1) the number of the bombarding primary electrons, (2) the velocities of the primary electrons, (3) the type of material used for the bombarded surface, (4) the physical condition of the surface and the state of the vacuum in which the emission occurs, and (5) the angle at which the bombarding primary electrons impinge on the bombarded surface. It is said that secondary emission is nearly independent of temperature, except that the high temperature may change the nature of the surface by changing the structure or by releasing absorbed gas. When primary electrons strike a surface at right angles, the secondary electrons are emitted at all angles and the spray of secondary electrons seems to nearly follow a cosine law of distribution. When the primary electrons strike a metal surface at other than a right angle, the distribution of the angle on the secondary electrons is still nearly a cosine-law variation. It is interesting to note that the secondary-to-primary emission ratio increases as the primary electrons strike more nearly parallel to the surface. Insulators as well as conductors may emit secondary electrons and the physical behavior observed is quite similar to that of metals.

Photoelectric Emission

Photoelectric emission of electrons occurs when radiant energy in the form of electromagnetic waves strikes the surface of certain materials. These electromagnetic waves are usually those in the general region of light radiation. Although early research in photoelectric emission phenomena was performed in

connection with metal surfaces, modern devices employ semiconducting materials as emissive surfaces almost exclusively. It is believed that photoelectric emission is largely a surface phenomenon and that radiant impinging energy of the proper wave length is able to impart sufficient energy to the free electrons near and at the surface and thereby to enable their escape.

Although the exact nature of photoemission is obscure (as are the exact natures of most fundamental phenomena), several laws are known and will now be reviewed.

1. The number of photoelectrons (electrons released by radiant energy) emitted varies directly as the intensity of the light (or other radiation) striking the photosensitive surface.

2. The initial velocity of the photoelectrons released from a metal is independent of the intensity of the light falling on the surface. A feeble beam of light from a distant star will release electrons with the same velocity as will a powerful light beam of the same wave length. (Of course, the light from the distant star will release electrons at a lower numerical rate as indicated above.)

3. The velocity of the ejected photoelectrons depends on the frequency (or wave length) of the light or other radiation releasing it. If the frequency is high enough to release electrons, then increasing the frequency (reducing wave length) will cause the electrons to be ejected with higher initial velocities.

4. Photoelectric emission is almost entirely independent of the temperature of the metallic surface.

Experiments have shown that photoelectron emitting materials exhibit a work function which corresponds very closely to that for thermionic emission as previously discussed. As the frequency of the impinging radiation is decreased, a lower frequency limit (long wave length) is reached at which the electrons cannot absorb sufficient energy to escape from their cathode. By a series of brilliant experiments, Millikan demonstrated conclusively that, for any given photoelectric material, there is a definite frequency of light below which no emission will take place no matter how great the intensity of the light or how long it may be applied. This point is known as the "threshold frequency."

Photoelectric emission has been the technical keystone in the development of camera and phototubes. More recently, practical photoconductive devices have been built that are quite competitive with camera tubes designed around photoemissive phenomena and phototubes. Photoelectric-emission phenomena have had the greatest importance in contributing to the establishment of contemporary theories pertaining to the basic mechanisms of physics. Planck's quantum theory was corroborated by Einstein on the basis of postulations involving the photoelectric effect. From these seemingly unrelated beginnings came the Einstein mathematical prognostications regarding the entire field of nuclear energy.

High-Field Emission

One more method of releasing electrons from metals relies on reducing the effectiveness of the attractive force of the metal ions near the surface of a cathode by the application of a strong external electric field. This phenomenon is known as high-field emission, a process whereby electrons are emitted from a cold solid or liquid because of a very high potential gradient. Such a gradient must be of the order of a million volts per centimeter, or higher, depending on the work function of the particular cathodic material employed. A high potential gradient does not necessarily imply high voltages since the gradient will depend upon the shape of the emitting material. Thus, if a fine metal point is dipped into chemicals to reduce it to a point of infinitesimal size, the application of several hundred volts can easily produce a potential gradient of a million volts per centimeter or more. This principle has been used in constructing at least one type of electron microscope. The currents extracted from metals by strong electric fields are substantially independent of temperature, until a point is reached where thermionic emission becomes appreciable. Theoretically, the high gradients partially neutralize the potential barrier across which emitted electrons must pass to escape from the cathode. High-field emission is also of importance inasmuch as it is frequently an unwanted mechanism in vacuum tubes because it is capable of freeing unwanted spurious electrons with deleterious effects on the performance of the tube.

The mercury pool is extensively employed as a source of electrons in various devices, e.g., the familiar mercury-arc and ignitron rectifier tubes. It is thought that the electrons are freed from such mercury pools by field emission but not from a fine point as previously considered. In these rectifiers, the strong electric field is produced by positive ions accumulating very close to the surface of the mercury pool. A similar action occurs in cold-cathode gas tubes.

Radioactive Disintegration

A radioactive substance is one whose atomic nuclei are unstable. Such an atom can exist for an indefinite period of time and then suddenly, for no apparent reason will explode, that is, the nucleus explodes. It shoots out an alpha particle (the nucleus of a helium atom) or a beta particle (electron). In the rearrangement of the residue, gamma rays (exceedingly short wave length electromagnetic radiations) also may emanate. In essence, there is a gradual decay of the radioactive element. The rate of decay, or disintegration, is independent of external influences such as heat and light, except for the influence of bombardment by other high-velocity nuclear particles. Certain elements also become artificially radioactive as a consequence of bombardment by high-velocity nuclear particles. The beta particles or electrons emitted by radioactive disintegration are similar in characteristics to electrons emitted by the previously mentioned mechanisms, save for the fact that they are emitted with an excess of energy and consequently they move at high velocities.

The high velocity of radioactively emitted electrons

has prevented their applications in practical electron tubes because the emitted beta particles (electrons) are virtually uncontrollable. However, scientists have been able to "catch" them; they have recovered and stored a portion of their kinetic energy to provide a source of low power for other electronic devices. The harnessing of wild emitted beta particles (electrons) is a fascinating endeavor which may produce practical devices in the future and experimentation is now being conducted with exceedingly low power devices of this sort.

Practical use has been made of the emitted alpha particle (helium atom nucleus). In the Alphatron vacuum gauge, a tiny radioactive emitter spews alpha particles that are capable of ionizing the residual gases in the vacuum system under observation. The residual ionization products are collected and metered to serve as an indication of the degree of vacuum in the vessel.

LAWS OF ELECTRONS IN MOTION

The motions[3] of electrons in vacuum tubes and other devices used in electronics are controlled by electric or magnetic fields, or both. Comprehension of the nature of these fields is perhaps one of the most rigorous exercises in mental gymnastics in the sciences. Practicing engineers develop a feeling for the reality of these fields in the physical sense of the word, visualizing these fields as being constituted of lines (or tubes) of force. Scientists, on the other hand, tend to think of these fields in a more abstract sense as mathematical concepts described by vector analysis. These fields are of paramount importance because they control the direction or the magnitude of electron flow, or both. In many applications of electronics, the electric fields are dominant in that the numbers of electrons involved is so small that they will not alter the applied field. In practical vacuum tube situations, an understanding of electron-flow behavior is complicated by the fact that electrons are usually present in numbers large enough to influence or distort the fields in consequence of their own "space-charge." For purposes of discussion, the motion of electrons as influenced by fields can be illustrated in three categories: (1) electron motion in a uniform electric field, (2) electron motion in a uniform magnetic field, and (3) electron motion in combined electric and magnetic fields.

Electron Motion in a Uniform Electric Field

An electron behaves like a small negatively charged particle and is therefore influenced by an electric field. Thus, suppose that an electron finds itself between two parallel-plate electrodes (see Fig. 1) and that the electrodes are in an evacuated tube. If a voltage is applied to the two electrodes to make one positive and the other negative as indicated, an electric field will be established between the electrodes approximately as indicated by the arrows. Thus, if an electron in an electric field moves freely toward a positive electrode, it has work done on it by charges on the electrode; it will gather speed and its kinetic energy will be increased. It should be understood that the laws of motion of an electron in a uniform electric field are the same as those of a body falling freely under the influence of gravity. If an electron is caused to move against the force of an electric field (for instance, if it were shot toward a negative electrode) the electron will do work on the electric charges on the electrode, will lose kinetic energy, and will slow down.

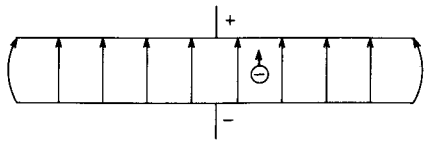


Figure 1. The Electron in an Electric Field Produced by Two Electrodes

An example of two-dimensional electron motion is shown in Fig. 2. Suppose that an electron enters between two parallel-plate electrodes between which a uniform electric field is established in the y-direction shown. The electron has an initial velocity v_{ox} in the $^{+}$x-direction. Since there is no force in the z-direction, there can be no acceleration component in that direction. Also, since the initial velocity in the z-direction is assumed to be zero, the motion must take place entirely in one plane, the plane of the paper. Likewise, there is zero force acting along the x-axis which means that the electron velocity along the x-axis remains constant and equal to the initial velocity v_{ox}. On the other hand, a constant acceleration exists along the y-direction such that the electron is accelerated upward, the velocity component v_y varies from point to point, whereas the velocity component v_x remains unchanged in the passage of the electron between the plates. It can be shown analytically that the electron moves in a parabolic path in the region between the plates.

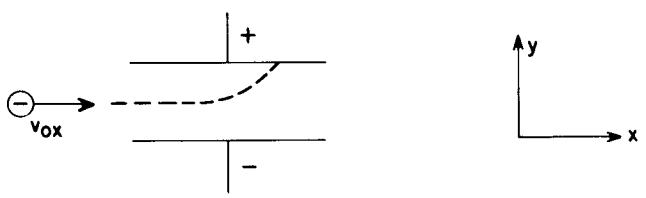


Figure 2. Two-Dimensional Motion in a Uniform Electric Field

Electron Motion in a Uniform Magnetic Field

An electron in motion constitutes an elementary flow of electricity and, as such, establishes an external magnetic field. This magnetic field will react with other magnetic fields and will produce a force which acts on the electron. This force is similar to the force acting on a wire carrying a current in a magnetic field. An electron has no external magnetic field when it is at rest and, thus, stationary electrons have no magnetic forces exerted on them when they are situated in a magnetic field. It is important to note the difference between the action of an electric and a magnetic field in this respect.

The force exerted on a wire carrying a current in a magnetic field is given by the following relation:

$$f = \frac{B\,I\,l}{10} \qquad (2)$$

where f = the force in dynes
I = the current in the wire in amperes
B = the magnetic induction in gausses (magnetic lines of force per square centimeter) at right angles to the wire
l = the length of the wire in centimeters

The force f of Eq. (2) is the result of the reaction between the magnetic field caused by the current flowing in the wire, and the externally produced magnetic field in which the wire is immersed. The current of moving electrons really produces the magnetic field encircling the wire. Thus, it follows that a force should be exerted on electrons moving at right angles through a magnetic field, whether these electrons are in a wire or moving as a stream through space. Experiments have shown this conclusion to be true. Again, referring to Eq. (2), it is known from fundamental electrical theory that the force exerted on the wire is at right angles to the wire carrying the electron current along it. It follows, therefore, that the force exerted on a stream of moving electrons is at right angles to the direction of the moving stream, and that the force exerted on an individual electron at each instant is at right angles to the direction of motion of the electron at that instant. This principle can be applied in determining the path which a moving electron will follow when the electron is in a magnetic field. There are two conditions to consider: (1) a moving electron being shot from outside into a magnetic field, and (2) a moving electron being released inside the magnetic field.

The first of these conditions is illustrated by Fig. 3 which shows an electron being shot at right angles into a limitless, constant, uniform magnetic field. At each instant, the force exerted on the electron by the magnetic field is at right angles to the direction of motion at that instant, and the path will be circular as indicated. The electron obeys laws which are kindred to the laws of circular motion encountered in the mechanics of particle motion. An examination of Fig. 3 shows that an electron shot into the field from a point external to the field will be forced out on the same side from which it entered if the field is limitless in the directions shown. If the field is finite, the electron will be deflected and will be forced out at the distant side, or out at the top, if it does not make a complete reversal in direction. In other words, in the case of electrons in magnetic field, a moving electron that enters from outside at right angles to a magnetic field, or in any direction not parallel to the lines of force, will be forced out of the field.

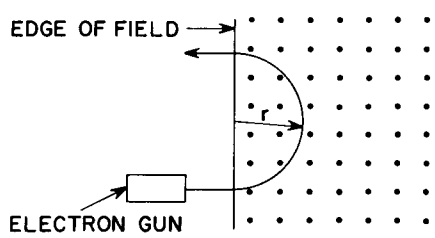


Figure 3. Path of Electron Shot Into a Magnetic Field

The second condition to be considered is shown in Fig. 4. An electron gun situated in a limitless uniform magnetic field shoots an electron in the field at right angles to the direction of the magnetic lines of force. Under these conditions, the electron must again assume a circular path. If the field is not limitless, the electron may, of course, leave the field, depending on its velocity, the field strength, and its original position in the field.

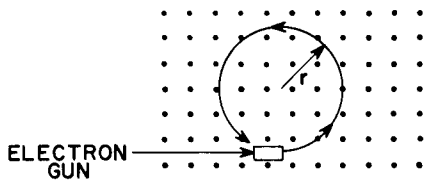


Figure 4. Path of Electron Emitted in a Limitless Magnetic Field

Irrespective of the shape of the path or other details, it is important to note that the force exerted by a magnetic field on an electron in motion is always at right angles to the direction of motion. Thus, a constant magnetic field can do no work on an electron because the electron will have the same kinetic energy when it leaves a magnetic field as when it enters; however, its direction of motion may be altered. This effect is in contrast with the action of an electric field which can increase or decrease the energy of an electron by altering its velocity as previously described.

As a corollary to the foregoing discussion it is fruitful to consider the influence of a changing magnetic field on an electron, particularly in view of the observation that a constant magnetic field can do no work on an electron although it can alter the direction of electron motion. The betatron[4] is a machine employed by physicists to accelerate electrons to exceedingly high velocities; it is a classical example of the forces associated with a changing magnetic field. In the operation of the betatron, electrons are injected into a circular, doughnut-shaped tube. When a changing magnetic field is applied parallel to the axis of the doughnut-shaped tube, two effects are produced: first, a radial force due to the action of the magnetic field, which is perpendicular to the electron velocity, keeps the electron moving in a circular path as has been depicted in Fig. 4. Second, an electromotive force tangential to the electron orbit is produced by the changing magnetic flux and gives the electron additional energy. This latter accelerating force is easily explained by thinking of the circular electron path of fixed radius as a circuit, and the emf (V) induced in this circuit results from the changing magnetic flux in accordance with Faraday's law:

$$V = \frac{d\phi}{dt} \qquad (3)$$

where ϕ = the instantaneous value of the magnetic flux perpendicular to the plane of the circuit.

Thus, in the betatron, the electrons acquire their increased energy by being accelerated around the dough-

nut several hundred thousand revolutions in a circular orbit as a consequence of the forces exerted on them by the electric field which accompanies a changing magnetic field.

Electron Motion in Combined Electric and Magnetic Fields

When an electron is subjected to the combined action of both electric and magnetic fields, the paths tend to become quite complex. Three of the simpler cases will be considered.

When an electron starts from rest under the influence of parallel electric and magnetic fields, the electron moves in the direction of the electric field and is unaffected by the magnetic field. The path in this case is a straight line, and the electron behaves as though the magnetic field did not exist.

If an electron with a given velocity is injected into a region containing electric and magnetic fields at right angles to each other and at right angles to the initial velocity, then there is a certain ratio of electric-field strength to magnetic-field strength for which the electron is not deflected in its path. This deflection occurs when the force due to the electric field is equal and opposite to that produced by the magnetic field.

Fig. 5 illustrates the case in which an electron starts from rest in the presence of uniform electric and magnetic fields that are mutually perpendicular. The electric field alone would direct the moving negative electron upward, and the magnetic field alone would cause the path to be circular. The result is that the moving electron is acted on by two vector forces; the direction of the electron's path at any instant will be determined by the vector sum of these two forces. Complex cycloids and trochoids of motion result in response to the permutations and combinations of fields which may be established.

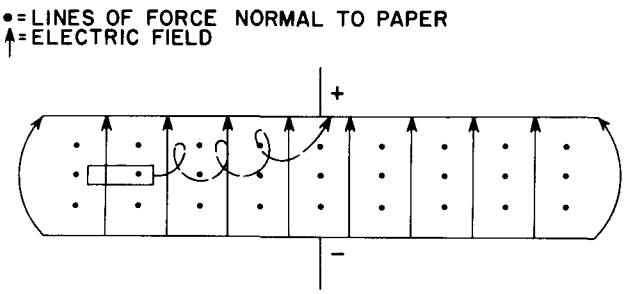


Figure 5. Path of an Electron Shot Into a Combined Magnetic and Electric Field

ELECTRON OPTICS[5, 6, 7]

As the term electron optics implies, there is a close analogy between the behavior of light rays and electron beams. This analogy not only has mathematical significance, but is also applicable to practical applications in the reflection, refraction, and focusing of electrons. As geometrical optics is concerned with the paths of light rays, electron optics deals with the determination and control of electron trajectories.

The importance of electron optics is becoming increasingly apparent with the continual advances in electronics. For example, in the early vacuum tubes used in radio work, little attention was given to the exact paths of the electrons between the cathode and plate. Today the design of virtually every modern vacuum tube is predicated on at least one electron-optical principle.

Electrostatic Electron Optics

The analogy between the behavior of light rays and electron beams is particularly close when the fields through which the electron moves are purely electrostatic. Electrons move through an electric field just as do light rays through a medium of continuously variable index of refraction.

The basic law of geometrical optics is Snell's law of refraction, from which all the properties of physical lenses can be deduced. This law has its exact counterpart in electron optics. Snell's law for optics is

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \qquad (4)$$

where n_1 and n_2 are indices of refraction on two sides of a plane boundary
θ_1 and θ_2 are the angles of incidence and refraction of a light ray as measured from a normal to the boundary

The corresponding situation for electron optics is shown in Fig. 6. This figure shows the behavior of an electron moving in a region with a uniform potential E_1 and suddenly crossing into a region with a uniform potential E_2. In traveling from the region of one potential to the other, the component of velocity normal to the boundary is increased if the potential is increased, but the tangential component of velocity is unchanged. Equating the initial and final tangential components of velocity gives

$$v_1 \sin \theta_1 = v_2 \sin \theta_2 \qquad (5)$$

where $v = 5.93 \times 10^5 \sqrt{E}$ (meters per second) when the potential (E) is given in volts.

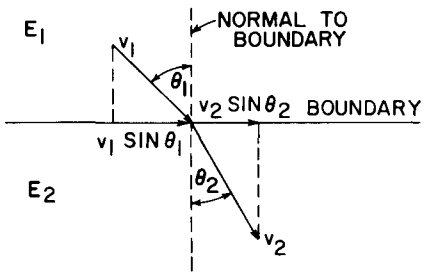


Figure 6. Electron Refraction

Comparing Eqs. (4) and (5), it is seen that the quantity in electron optics corresponding to the index of refrac-

tion is electron velocity. It should be clearly understood that electron velocity does not correspond to the velocity of light.

In the idealized analytical design of electrostatic electron-optical lenses, it should be possible to obtain mathematical expressions for the potential fields associated with any given set of electrodes and then to solve for the path of an electron through these fields. Actually, the fields of electron lenses are not necessarily simple to determine and the solution to electron trajectories through them is even more complex. Experimental and analytical determinations of lens characteristics have been complementary in the development of scientific techniques for the design of electron-optical systems. The application of purely electrostatic electron-optical principles to vacuum tubes is exemplified in beam power tubes[8] and electrostatically focused traveling-wave tubes (Estiatrons).[9]

Magnetic Lenses

Electron beams can be focused with magnetic fields as well as with electric fields, though the analogy with optics is not readily apparent. It has already been pointed out that an electron moving in a magnetic field will not be affected by the field if its motion is parallel to the field. However, if it is moving at an angle with respect to the lines of magnetic flux, it will be forced to execute a helical path with its axis of motion parallel to the lines of force. For example, consider an electron starting from a point O with a linear velocity V and making an angle α with the flux lines produced by a long solenoid. The component of motion along the field will not be affected, but its component, V sin α, at right angles to the field, will be acted on by a force at right angles both to the direction of motion and to the magnetic field. Thus, the component V sin α will be transformed from a straight line to a circle. According to the equation for the motion of an electron in a magnetic field,

$$\frac{mv^2}{r} = Hev \tag{6}$$

where m = mass of the electron
v = V sin α (where V is linear velocity)
r = radius of orbit
H = field strength
e = electronic charge

Hence,

$$r = \frac{mv^2}{Hev} = \frac{v}{\frac{e}{m}H} \tag{7}$$

and the time required for one revolution is

$$t = \frac{2\pi r}{v} = \frac{2\pi}{\frac{e}{m}H} \tag{8}$$

which is independent of both V and α.

This circular motion combined with its linear motion in the direction of the field causes the electron to traverse a helical path with its axis parallel to the field. Thus, all the electrons starting from a given point O will arrive at a point O' after a time interval t. In this way, an exact erect image of O can be formed at O'. In practice, a short focusing coil is used instead of a long uniform field. This coil is quite analogous to an optical lens and the size of the electron image is in accordance with the optical-lens formula.

$$\frac{1}{O} + \frac{1}{i} = \frac{1}{f} \tag{9}$$

where O = object distance
i = image distance
f = focal length

A study of the electron microscope is suggested as an exemplary application of magnetic lenses to electron beams.[10]

Periodic Focusing

In some of the simplest electron-optical systems, the focusing fields vary periodically along the electron stream and constitute a series of thick or thin electron lenses. Electric focusing of this sort was used in some of the earliest velocity-modulated tubes.[11] During the past decade, there have been spectacular advances in the utilization of periodic focusing systems, both electric and magnetic. In 1952, Courant, Livingston, and Snyder[12] suggested that focusing may be accomplished by a series of alternating converging and diverging lenses of equal strengths arranged periodically. This "strong-focusing" or "alternating-gradient focusing" principle revolutionized the design of High-Energy Particle Accelerators and inspired new investigations of periodic focusing. In 1953, Pierce[13] amplified the theory of periodic focusing for low-voltage electron beams with additional contributions by Clogston and Heffner.[14] A practical traveling-wave tube employing periodic magnetic focusing was reported[15] by Mendel, Quate, and Yocom in 1953.

Meanwhile, research in electrostatic periodic focusing was equally fruitful. In 1957, Chang[16] reported the successful operation of an experimental traveling-wave tube in which an electrostatically confined electron flow was obtained by utilizing the centrifugal force of rotating electrons as a restoring force to balance the strong focusing force due to a periodic electrostatic field of very short period. Chang[17] reported a second experimental tube demonstrating the superiority of two counteracting periodic fields of very short periods as established by a bifilar helix in the rf interaction region. The potential valley formed by the combination of two counteracting periodic fields is steeper than all previously reported focusing systems, thereby maintaining a very stable electron flow in the beam. Furthermore, in this biperiodic focusing system, the force balance on the beam electron in the confined-flow condition is independent of the average potential. In the drift region of traveling-wave tubes where electrons interact with the rf delay line, if the beam focusing is not dependent upon the dc velocity of electrons, defocusing caused by rf interaction may be avoided. Traveling-wave tubes employing all-electrostatic beam generation and focus are called Estiatrons,[9] after the Greek word ESTIA meaning to focus.

What are the advantages of using periodically vary-

ing magnetic fields rather than magnetic or electric fields which do not vary with distance? In the case of electric fields, it can be said that such focusing provides one method of avoiding magnetic fields with attendant reductions in weight. The use of periodic electric fields may be practical in some tubes because the electrodes required either do not interfere with other necessary electrodes or perhaps may actually be combined with other electrodes. In other cases, it may be impractical or undesirable to provide the electrodes necessary to produce periodic electric focusing fields along the beam.

There are at least two major advantages in employing periodic magnetic focusing fields instead of long solenoids with uniform magnetic fields. First, the concentration of the magnetic field into short regions near critical gaps can increase the focusing action. Second, from a practical standpoint, the effectiveness of periodic focusing in a particular tube may be such that its magnet weight may be a tenth or less than that required to produce a uniform field with the same focusing effectiveness.

Electron Guns

Electron-optical principles have found their widest application in the development of electron guns, an electronic system comprising a cathode together with electrodes for the formation, control, and focusing of electron beams. The development of cathode-ray tubes served as an impetus to the evolution of the electron-gun art; examples of contemporary designs are described by Moodey.[18] Thompson and Headrick[19] described basic calculations for electron-gun performance in conjunction with an electrostatically focused beam in a cylindrical tunnel. Smith and Hartman[20] performed calculations for electron-gun performance with electron beams in tunnels, in which beam spreading was prevented by the application of a steady magnetic field parallel to the beam axis. In his classic paper, Pierce[21] outlined the principles of the "Pierce Gun" that has been utilized in virtually all electron guns for klystrons and traveling-wave tubes. Samuel[22] and Pierce[23] have written notes on the design of these guns. Bliss[24] has described modern examples of electron guns incorporated in traveling-wave-tube designs.

The presence of beams in grid-controlled tubes is not new; electron beams were formed by the electrostatic fields of the grid in the DeForest triode. More recent knowledge of beams and their properties, their use, and application is described by Thompson[25] and by Schade.[26] Performance characteristics of power tubes have been improved spectacularly by the incorporation of electron-beam technology. Super-power tubes, described by Bennett,[8] employ unitized electron guns in which electron beam-forming is accomplished by the application of electron-optical principles to the design of cathode and grid structures. The inductive-output[27] tube is another illustrative example of electron-gun and electron-optical technology.

ELECTRON SPACE CHARGE

The presence of electrons in the interelectrode spaces of vacuum tubes can be responsible for a number of very important effects in the operation of tubes. A synoptic review of these effects follows and gives pertinent literature references in which comprehensive information may be found.

Space-Charge Effects in Diodes

In a diode, the electrons may be visualized as negative charges in space and so constitute a negative space charge in the region between the cathode and plate. As a consequence of this space charge, the total accelerating force on electrons near a positive plate is greater than that on the electrons near the negative cathode. The electrons near the plate accelerate faster than those near the cathode and the electron concentration or density thins out near the positive plate as shown in Fig. 7. In summary, the influence of negative space charge greatly alters the distribution of the voltage applied between the cathode and plate of a diode and affects the voltage-vs.-current characteristics of a diode. This effect and its relationship to temperature and voltage saturation conditions have been discussed in detail by Albert,[3] Spangenberg,[2] and Fink.[28]

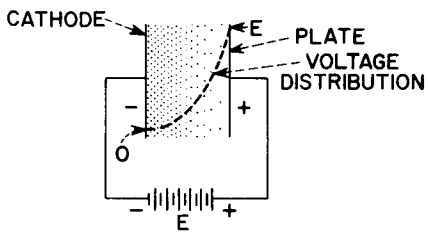


Figure 7. Effect of Voltage on Electron Distribution

Space-Charge Effects in Triodes

If a grid is situated between the negative cathode and the positive plate and if this grid is connected to a battery of E_c volts while the plate is connected to a battery of E_b volts, then the voltage distribution between the plate and cathode depends not only on the plate voltage but on the grid voltage as well. If the grid is made positive with respect to the cathode by a sufficient amount, then the voltage-distribution curve is "lifted" by the grid as shown in Fig. 8A. If the grid is made negative with respect to the cathode, then the voltage-distribution curve is depressed, as in Fig. 8B. This lifting or depressing of the curve changes the slope of the voltage-distribution curve at the cathode surface and, hence, changes the force acting on the electrons (the force is proportional to the slope). When the grid is positive, the force is great; but when the grid is negative, the force is decreased; it may be decreased to zero if the voltage on the grid is sufficiently negative. Under these conditions, the electrons are not urged away from the emitting cathode at all, and current ceases entirely as depicted in Fig. 8C. Moreover, if the grid is negative, it cannot collect electrons, and, hence, it can control the electrons without interfering with the current itself. This fact makes grid control a highly efficient and sensitive method of changing electron current. In essence, a space-charge control-grid in a triode functions in this manner. The references by Albert,[3] Spangenberg,[2] and Fink[28] treat this subject comprehensively.

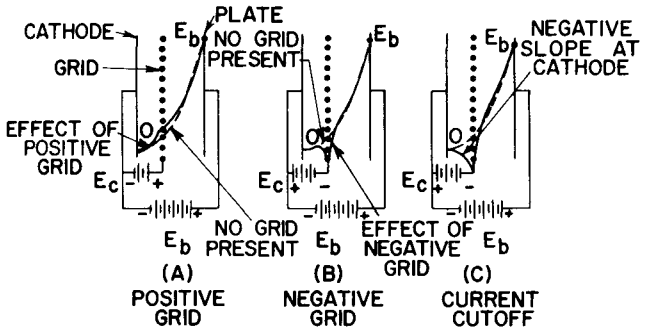


Figure 8. Effect of Grid Voltage on Voltage Distribution

Effect of Space Charge on Electron Transit-Time

Although electron velocities in vacuum tubes are comparatively high, the time required for the electrons to pass from one electrode to another is appreciable. At low frequencies, the time required for the passage of an electron from the cathode through the grid and to the plate is of little importance. At much higher frequencies, this statement may no longer be true. The time required for an electron to pass from the cathode to the plate, called the "transit time," may be comparable with the time required for one frequency cycle to occur at the much higher frequency. Thus, an electron leaving the cathode under a given phase relationship may find that the phase conditions have changed markedly by the time it reaches the vicinity of the plate. This effect is of vital importance in tubes used at microwave frequencies.

In the preceding sections it has been pointed out (as shown in Figs. 7 and 8) that the presence of electrons between electrodes causes the effective accelerating potential to be lower than it would be otherwise. Consequently, the electron velocity is lower than it would be otherwise and the transit-time of electrons through the tube is significantly greater. In diodes, for example, it is not unusual for the transit-time to be increased threefold due to increased space-charge effects. The entire situation becomes almost incomprehensibly complex from an analytical standpoint under conditions in which the accelerating electric field is varying with time. An appreciation of the complexity of these problems can be gleaned from the fact that in diodes the space charge limits the currents to a value determined by the three-halves power of plate voltage and also causes the potential distribution within the tube to be nonlinear, as shown in Fig. 7. The combination of these nonlinear effects greatly complicates analytical approaches to solutions for transit-time problems in the situations wherein the accelerating voltage is varying with respect to time at a high radio-frequency rate. Kleen[29] has published expressions and curves for the determination of transit-time taking space charge into account.

Space Charge and the "Virtual" Cathode

In the presence of space-charge saturation, electrons having zero velocity would never get started from a cathode. Fortunately, electrons are emitted from a cathode with finite initial velocities, about 90 per cent of them having velocities (at usual cathode temperatures) equivalent to an accelerating potential of less than one-half volt. The presence of electron space charge in the proximity of the cathode produces retarding fields and slows down the emitted electrons until they all momentarily come to rest at a potential minimum. From this region, which acts like an ideal cathode and is called a "virtual" cathode, the electrons may start in either direction, either being returned to the cathode or proceeding away from the cathode to the plate. The distance of the virtual cathode from the actual cathode may be appreciable. For a cathode temperature of 1000 K and a transmitted current density of one milliampere per cm^2, the distance from the cathode to the virtual cathode is approximately 0.006 inches. In modern close-spaced electrode tubes, this distance may be effectively very large.

The consideration of virtual cathode effects is of importance in at least two other cases in vacuum tubes. In beam power tubes, for example, a region of high space-charge density and low potential is deliberately produced in the region between screen grid and plate by means of juxtaposed beam-forming or zero-potential deflecting plates.[30,31] This region of lowered potential performs the same function as the suppressor grid in a pentode, i.e., it prevents the backward flow of secondary-emission electrons from the plate to the screen grid. Virtual cathodes are also encountered in electron-optical systems, e.g., in cathode-ray-tube systems.[1]

Space Charge and Electron-Beam Dispersion

If an electron beam is not confined by an axial magnetic field, it will tend to spread because of space-charge repulsive forces. An electron on the outer surface of the beam experiences an acceleration due to the radial electric field which originates from the charge in the beam. In devices requiring a beam of fairly high electron-current density, such as those used in cathode-ray tubes, and especially in microwave amplifying tubes of the beam type, the fields caused by the charge in the beam itself cause serious difficulties with divergency and limit the electron-current densities which can be handled. Even with the restraining forces of axial magnetic fields, practical experimentation has indicated limits on electron-current densities in beam transmission. Brillouin[32] has published solutions for electron motion in magnetically confined cylindrical beams, in hollow beams, and in two-dimensional beams. Pierce[23] has derived practical design equations for the effects of space charge in electron beams.

In cathode-ray tubes, the production of small beam spots is seriously limited by the space-charge mutual repulsion between electrons in the beam, thereby preventing the electrons from coming together to give sharp beam focus.[19] In a convergent beam, as the beam tends to come to a smaller diameter, the electrons get closer together, the space-charge density increases within the beam, and hence the mutual-repulsion forces become greater.

Space-Charge Waves[33, 34]

It has been shown previously that space-charge effects can appreciably influence the characteristics of vacuum tubes and electron beams. In essence, the classical grid-controlled tube is an electron-discharge device in which the control-grid voltage disturbs the space-charge region and functions as a "valve" on flow of electronic current through the tube. Used in appropriate circuits, these types of devices serve as amplifiers, oscillators, frequency converters, modulators, detectors, electronic computing devices, special wave form generators, control devices, and measuring devices. In general, these grid-controlled devices produce external-circuit currents which vary in accordance with the number of electrons striking one of the electrodes. In the parlance of the art, electron streams in grid-controlled tubes are said to be "density-modulated."

The invention of the klystron by the Varian brothers[35] marked the introduction of an entirely new class of devices in which the electron stream is depicted as being "velocity-modulated." At about the same time, Hahn[36] and Ramo[37] were investigating the propagation of waves in the medium represented by an electron stream. These waves are called, characteristically, space-charge waves. In the following paragraphs the "velocity-modulation" of a klystron will serve as an illustrative model and example in the more comprehensive field of space-charge waves.

Let it be imagined that a beam of electrons has been accelerated and caused to flow into the left-hand side of Fig. 9 at a uniform velocity. So long as the electrons enter this beam at a uniform rate and proceed in an orderly fashion, the average motion of the electrons is practically unaffected by interelectronic forces. There would, however, be a tendency toward space-charge enlargement of the beam cross section by dispersion unless a sufficiently strong magnetic field were arranged parallel to the z-axis to restrain the radial-force effects. The delicate balance of space-charge uniformity in the beam can be disturbed by some influence such as an electric field ($V_1 e^{j\omega t}$) at a certain position (z equals zero) established by a pair of grids, between which appears an alternating voltage of frequency ω. It should be understood that the electric field is capable of imparting a sinusoidal disturbance such that, for example, when a group of electrons is slowed down, other electrons behind them crowd together while those electrons ahead are more sparsely distributed. The crowding together causes this section of the beam to be more negative in potential; the crowded electrons will find themselves urged toward the less-crowded region. Thus, a wave disturbance is launched in the space-charge region and is then propagated along the moving beam.

The situation just described is reminiscent of the action of molecules in a gas, where the particles in a region of condensation tend to move into a region of rarefaction and the resultant disturbance is propagated through the gas as a sound wave. A condensation or rarefaction of electron density propagates down the beam with a characteristic velocity analogous to the

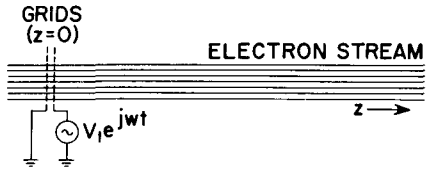


Figure 9. Model for the Analysis of Space-Charge Waves

velocity of sound in a gas. One must be careful, however, not to take this analogy between sound waves and space-charge waves too seriously because the phenomena are fundamentally quite different. In the sound-wave case, neighboring molecules exert an influence through elastic collisions which transfer kinetic energy from one molecule to another. In the case of space-charge waves, the electrons are charged and exert an effect on one another through Coulomb's-law forces, such that an individual electron is influenced constantly by the positions of all other electrons.

Collisions, in the usual sense, do not occur between electrons. The velocity of a group of electrons, as seen by a stationary observer viewing Fig. 10, will appear to have a constant value with a superimposed alternating or undulatory component. Similarly, the electron density will have a constant average value plus superimposed fluctuations. If this perturbed beam of electrons is permitted to continue drifting, eventually some of the faster electrons will move up on slower electrons which left the grids earlier to start the formation of "bunches" of electrons. At the same time, these concentrations of space charge have set up electric fields in such a direction as to cause the "bunches" to fly apart, and these same electric fields have reduced the velocity fluctuations to zero. At this juncture, current fluctuations in the stream have reached a maximum amplitude and a second set of grids should be placed across the stream to "catch" undulations of space-charge density in the beam and convert them into electrical energy in an external circuit. Fig. 10 depicts the build-up of charge "bunches" prior to their passage through "catcher" grids which thereby induce currents in the external circuit. The periodicity of these charge "bunches" is dependent upon the velocity of space-charge wave propagation through the medium, i.e., the electron beam. Propagation, in turn, is dependent upon the characteristics of the medium, which is determined entirely by the current density of the electron beam. This periodicity of the space charge in the medium is termed the plasma wave length. It should be appreciated that this explanation has been an abbreviated description of complex processes. Comprehensive descriptions are given in the literature by Beck,[33] Harman,[34] and Beck.[35]

In summary, it should be realized that the effectiveness of coupling electron-beam energy to radio-frequency circuitry (e.g., cavity resonators) is contingent upon the establishment of maximized density-modulated pulses of space charge in the electron stream. This requisite is established at the outset in grid-controlled tubes between the cathode and grid electrodes as "density-modulation." In space-charge wave devices (e.g., the klystron) an accelerated beam is "velocity-modu-

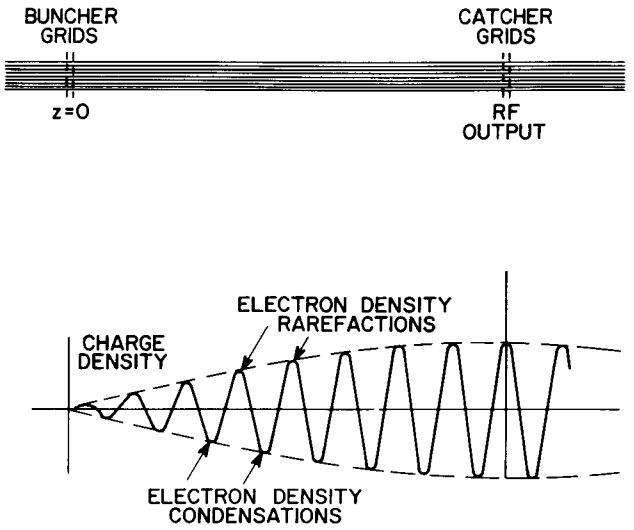


Figure 10. Model Depicting Build-Up of Space-Charge Waves

lated" and allowed to drift until a space-charge wave builds up "density-modulated" pulses for conversion into radio-frequency energy.

INTERACTIONS OF ELECTRONS WITH CIRCUITS[38]

Exchange of Energy Between Electron Streams and Electric Fields

Exchange of energy or power between an electron stream and circuit elements is always produced by accelerating or retarding the electrons by means of an electric field. These electron-to-circuit interactions are particularly important in microwave tubes wherein it is difficult to separate the tubes from their circuits. In microwave tubes, the kinetic or potential energy of electrons is converted into electromagnetic energy. Thus, klystrons[35] and traveling-wave tubes[39] are examples of devices in which kinetic electron-beam energy is converted into rf energy. In a crossed-field type of amplifier tube like the magnetron,[40] the potential energy of electrons is converted into rf energy in the cathode-anode region.

Circuit Currents Induced by Electron Motion

In many practical applications of electrons in vacuum tubes, moving electrons interact with circuits by the mechanism of induction. This type of electron-to-circuit interaction can be visualized by considering a single charge -e as it passes through a static electric field as shown in Fig. 11. The lines of force emerging from the charge terminate on the electrodes; the number of lines to each electrode are dependent on its distance from the charge. If the electron in Fig. 11 is moving from 1 to 2, the number of lines ending on electrode 1 decreases with time, while the number ending on electrode 2 increases. The induced charges on the electrodes vary similarly. In other words, as the moving electron approaches the second electrode, another electron is being pushed off electrode 2 to make room for the one approaching and thereby to maintain a balance of electronic charge. Simultaneously, still another electron is being forced to move into electrode 1. These charges in motion constitute a "series" flow of current, i.e., an electron is flowing between electrodes in a vacuum and forcing the flow of a "companion" or "image" electron in the external circuit. If the electron moving in vacuum has kinetic energy when it strikes electrode 2, this energy is dissipated as heat by increasing the molecular agitation of the electrode. The net result of an electron's passing from one electrode to the other is a total transfer of one electronic charge. At low frequencies, this fact is the only matter of concern. At high frequencies, when the electric field may change appreciably during the time of transit, the manner in which the moving charge induces a current in the external circuit is of major importance. This induced current should not be confused with displacement currents which flow when an alternating potential is impressed across a pair of capacitor plates. The induced current contributes to the usual capacitor displacement current an additional current component that is due solely to the motion of the electron in vacuum between two electrodes.

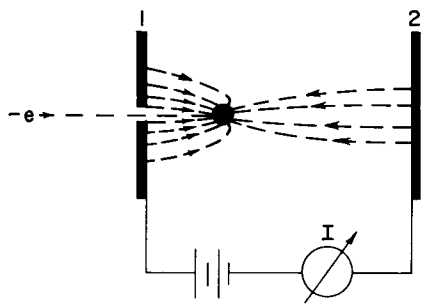


Figure 11. Electron-Induced Current in a Two-Electrode System

The induction of currents in circuits can be further illustrated by considering a plane-parallel three-electrode system such as that shown in Fig. 12. When an electron enters region 1-2 (point A) at a small aperture, an induced current begins to flow. This induced current $I_{ind,\ 1\text{-}2}$ flows until the electron reaches point B, when it enters region 2-3 and a similar induced current $I_{ind,\ 2\text{-}3}$ begins to flow. For very low frequencies, the current induced by a stream of electrons is equal to the convection current and the currents in M_1 and M_2 are equal if there is no electron interception at aperture B. At high frequencies, however, as a result of the skin effect, separate induced currents (and similarly separate capacitive charging currents) flow on the two surfaces of electrode 2. It should be understood that the current flowing on the left-hand surface of electrode 2 is similar to that flowing on the left hand surface of electrode 3 despite the fact that the electron convection current is not intercepted by electrode 2 as is the case for electrode 3. It is concluded from this analysis that in any system containing several regions separated by electrodes, one must consider the current flowing in each region separately, whether electrons are collected or not.

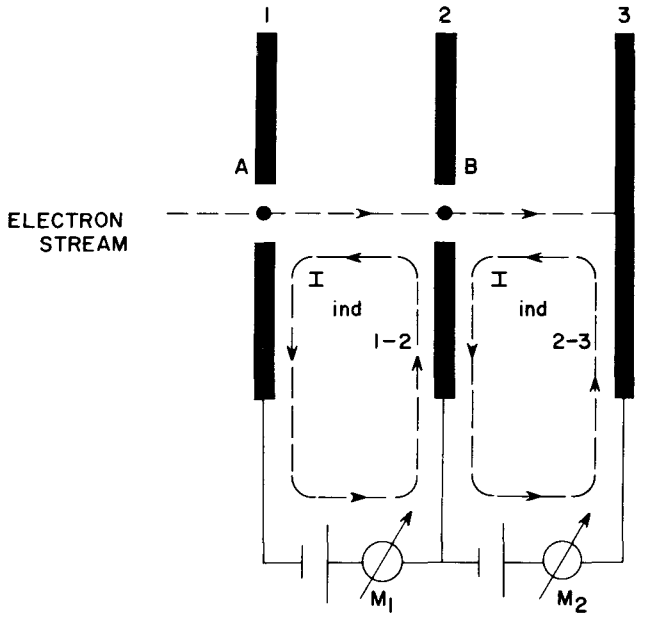


Figure 12. Electron-Induced Currents in a Three-Electrode System

Electron-to-Circuit Energy-Exchange Systems

There are several different types of systems whereby electron-to-circuit energy exchanges may be accomplished. The most common method of electron-to-circuit interaction has previously been described in connection with Figs. 9 and 10, where the radio-frequency fields are of the so-called "standing-wave" type. It should be noted that the electron stream is moving relative to a "standing wave" of stationary voltage established in both the bunches and cathode circuits. If these electron-to-circuit interactions are recurrent, the electrons are subjected to periodic rf fields. Thus, the reaction described in connection with Fig. 9 can be performed periodically with a multisectioned cascade arrangement of circuit resonator elements. Let Fig. 13 represent a cascaded group of cavity resonators traversed by an electron beam as is the case in a cascade klystron amplifier. The impressed voltage across the first resonator gaps produces a small amount of velocity modulation of the beam. By the mechanism of induction, a small amount of energy is extracted from the beam to excite the gaps in the second resonator thereby inducing additional bunching forces on subsequent electrons so that a greater alternating current is induced in the beam. The magnitudes of these effects increase along the length of the iterated structure. Thus, a modulation of the wave is produced which grows in amplitude as it progresses down the chain until finally it is limited by a saturation effect when the ac voltages and currents become comparable in magnitude to the corresponding dc quantities.

A second category of electron-to-circuit energy-exchange reactions involves a traveling wave of rf power. When an electron convection current flows through the field of a "traveling" radio-frequency wave, power is exchanged between the stream and the wave. Pierce[39] has described this type of electron-to-circuit interaction in connection with a traveling-wave tube in which the rf wave "travels" over a distance of 20 to 30 wave lengths. The electron velocity must be roughly equal to that of the wave. For non-relative velocities, the velocity u_o of an electron falling through V_o volts is

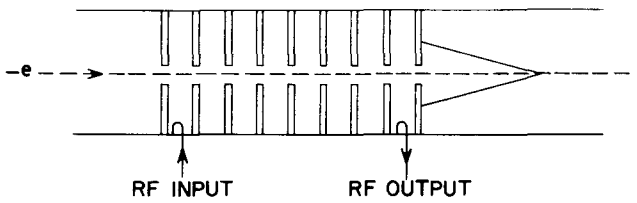


Figure 13. Cascaded Cavity-Resonators

$$u_o = 5.93 \times 10^5 \sqrt{V_o} \qquad (u_o \text{ in meters/second})$$

Therefore, if the electron beam is to interact with the "traveling-waves," the latter must be slowed down. Such waves are called slow waves and are established in juxtaposition to the electron stream by means of slow-wave structures, which also should ideally establish waves with a strong E_z component with respect to the beam. The helices commonly employed in traveling-wave tubes are the most common slow-wave structures currently in use for electron-to-circuit interaction.

Electron Beams in a Nonpropagating-Circuit Environment

Contemporary research and development efforts are revealing the attractive possibilities of employing electron beams in the environment of nonpropagating circuits. This class of microwave amplifier tubes depends on the behavior of space charge as a function of the dc characteristics of the beam (e.g., velocity, density) and the nature of the surroundings of the beam. The building block for these tubes is the electron beam itself in a nonpropagating environment. Examples of this class of tube are the "velocity-jump" amplifier,[41] the resistance-wall amplifier,[42] the rippled-wall amplifier,[43] and a space-charge wave amplifier with a varying beam diameter.[44, 45]

One of the most interesting examples of this class of tube is the electron-wave tube,[46, 47] also called the double-stream amplifier. This tube is a hybrid device that is a cross between space-charge-wave amplifier tubes and the classical traveling-wave tube. The electrons of two streams travel at slightly different velocities; preferably the two streams are thoroughly mixed (coupled together). This class of tube resembles a space-charge-wave amplifier tube in that the slow-wave structure is absent, and resembles a conventional traveling-wave tube (such as the helix-type tube) in that one of the two beams may be said to assume the role of the helix as a medium supporting a slow wave. Again, the building block is the stream itself, consisting of two beams. A kindred device, the "slipping-stream"[48] amplifier tube, is a device in which a magnetic field is applied perpendicular to the axis of beam motion.

REFERENCES

1. Millman, J., and S. Seeley, Electronics, McGraw-

Hill, New York, 1951
2. Spangenberg, K. R., Vacuum Tubes, McGraw-Hill, New York, 1948
3. Albert, A. L., Fundamental Electronics and Vacuum Tubes, Macmillan, New York, 1947
4. Kerst, D. W., "The Acceleration of Electrons by Magnetic Induction," Phys. Rev., Vol. 60, p. 47, July 1, 1941
5. Maloff, I. G., and D. W. Epstein, Electron Optics in Television, McGraw-Hill, New York, 1938
6. Klemperer, O., Electron Optics, Cambridge University Press, 1939
7. Cosslett, V. E., Introduction to Electron Optics, Clarendon Press, 1946
8. Bennett, W. P., "Large-Power-Tube Design, Processing, and Testing" (article in this book)
9. Blattner, D. J., F. E. Vaccaro, C. L. Cuccia, and W. C. Johnson, "Medium-Power L- and S-Band Electrostatically Focused Traveling Wave Tube," RCA Review, Vol. 20, p. 426, September 1959
10. Zworykin, V. K., et al, Electron Optics and the Electron Microscope, Wiley, New York, 1945
11. Hahn, W. C., and G. F. Metcalf, "Velocity-Modulated Tube," Proc. IRE, Vol. 27, p. 106, February 1939
12. Courant, E. D., M. S. Livingston, and H. S. Snyder, "The Strong-Focusing Synchrotron -- A New High Energy Accelerator," Phys. Rev., Vol. 88, p. 1190, December 1, 1952
13. Pierce, J. R., "Spatially-Alternating Magnetic Fields for Focusing Low Voltage Electron Beams," Jour. Appl. Phys., Vol. 24, p. 1247, Sept. 1953
14. Clogston, A. M., and H. Heffner, "Focusing of an Electron Beam by Periodic Fields," Jour. Appl. Phys., Vol. 25, p. 436, April 1954
15. Mendel, J. T., C. S. Quate, and W. H. Yocom, "Electron Beam Focusing with Periodic Permanent Magnet Fields," Proc. IRE, Vol. 42, p. 800, May 1954
16. Chang, K. K. N., "Confined Electron Flow in Periodic Electrostatic Fields of Very Short Periods," Proc. IRE, Vol. 45, p. 66, January 1957
17. Chang, K. K. N., "Biperiodic Electrostatic Focusing for High-Density Electron Beams," Proc. IRE, Vol. 45, p. 1522, November 1957
18. Moodey, H. C., and A. M. Morrell, "Color Picture Tube Design and Processing" (article in this book)
19. Thompson, B. J., and L. B. Headrick, "Space Charge Limitations on the Focus of Electron Beams," Proc. IRE, Vol. 28, p. 318, July 1940
20. Smith, L. P., and P. L. Hartman, "The Formation and Maintenance of Electron and Ion Beams," Jour. Appl. Phys., Vol. 11, p. 220, March 1940
21. Pierce, J. R., "Rectilinear Electron Flow in Beams," Jour. Appl. Phys., Vol. 11, p. 548, August 1940
22. Samuel, A. L., "Some Notes on the Design of Electron Guns," Proc. IRE, Vol. 33, p. 233, April 1945
23. Pierce, J. R., Theory and Design of Electron Beams, D. Van Nostrand, New York, 1954
24. Bliss, E., "Traveling Wave Tube Design and Construction" (article in this book)
25. Thompson, H. C., "Electron Beams and Their Applications in Low Voltage Devices," Proc. IRE, Vol. 24, p. 1276, October 1936
26. Schade, O. H., "Beam Power Tubes," Proc. IRE, Vol. 26, p. 137, February 1938
27. Haeff, A. V., and L. S. Nergaard, "A Wide-Band Inductive-Output Amplifier," Proc. IRE, Vol. 28, p. 126, March 1940
28. Fink, D. G., Engineering Electronics, McGraw-Hill, New York, 1938
29. Kleen, W. J., Electronics of Microwave Tubes, Academic Press, New York, 1958
30. Dreyer, J. F., "The Beam-Power Tube," Electronics, Vol. 9, April 1936
31. Haeff, A. V., and B. Salzberg, "Effect of Space-Charge in the Grid-Anode Region of Vacuum Tubes," RCA Review, Vol. 2, p. 336, 1938
32. Brillouin, L., "A Theorem of Larmor and Its Importance for Electrons in Magnetic Fields," Phys. Rev., Vol. 67, p. 260, 1945
33. Beck, A. H. W., "Space-Charge Waves," Pergamon Press, New York, 1958
34. Harman, W. W., Fundamentals of Electronic Motion, McGraw-Hill, New York, 1953
35. Beck, A. H. W., Velocity-Modulated Thermionic Tubes, Cambridge University Press, MacMillan, New York, 1948
36. Hahn, W. C., "Wave-Energy and Transconductance of Velocity-Modulated Electron Beams," General Electric Review, Vol. 42, p. 258, 1939
37. Ramo, S., "The Electronic-Wave Theory of Velocity-Modulation Tubes," Proc. IRE, Vol. 27, p. 757, December 1939
38. Hutter, R. G. E., Beam and Wave Electronics in Microwave Tubes, D. Van Nostrand, New York, 1960
39. Pierce, J. R., Traveling-Wave Tubes, D. Van Nostrand, New York, 1950
40. Collins, G. B., Microwave Magnetrons, Vol. 6 of the MIT Radiation Laboratory Series, McGraw-Hill, New York, 1948
41. Tien, P. K., and L. M. Field, "Space-Charge Waves in an Accelerated Electron Stream for Amplification of Microwave Signals," Proc. IRE, Vol. 40, p. 688, June 1952
42. Birdsall, C. K., G. R. Brewer, and A. V. Haeff, "The Resistive-Wall Amplifier," Proc. IRE, Vol. 41, p. 865, July 1953
43. Birdsall, C. K., "Rippled Wall and Rippled-Stream Amplifiers," Proc. IRE, Vol. 42, p. 1628, 1954
44. Peter, R. W., S. Bloom and J. A. Ruetz, "Space-Charge-Wave Amplification Along an Electron Beam by Periodic Change of the Beam Impedance," RCA Review, Vol. 15, p. 113, 1954
45. Mihran, T. G., "Scalloped-Beam Amplification," Jour. Appl. Phys., Vol. 25, p. 1341, October 1954
46. Nergaard, L. S., "Analysis of a Simple Model of a Two-Beam Growing-Wave Tube," RCA Review, Vol. 9, p. 585, 1948
47. Pierce, J. R., "Wave Picture of Microwave Tubes," Bell System Tech. Jour., Vol. 33, p. 1343, 1954
48. Guenard, P., and H. Huber, "Experimental Study of the Interaction of Space-Charge Waves in an Electron Beam Moving in Crossed-Electric and Magnetic Fields," Ann. Radioelectricite, Vol. 7, p. 252, 1952

Heaters

A. G. F. Dingwall

Harrison

Insulated heaters for electron tubes came into general use in the electronics industry in 1927 when ac power was first widely used for supplying heat to the cathodes of electron tubes. It was normally necessary then, as now, that the heater be isolated electrically from the cathode so that the ac power used in heating the cathode would not interfere with tube operation.

The commercial production of a satisfactory insulated heater presents many technical problems mainly because the high temperatures at which heaters customarily operate are well above the working limits of most materials. The trend has been towards the manufacture of smaller tubes requiring heaters which operate at higher temperatures; as a result, present-day heater requirements are considerably more stringent than those of twenty years ago. Heater failures and various heater troubles have always constituted a substantial proportion of the total causes of ultimate tube failure; consequently, careful attention to the design of heaters is necessary. Reliable heaters with satisfactory life and insulation characteristics can be obtained by proper design and processing and by the use of the best and most reliable types of materials available.

BASIC DESIGN CONSIDERATIONS AND THEORY

Heaters may be designed either by calculations based on theory or from a set of purely empirical nomographs prepared in 1947 by A. P. Kauzmann after an extensive survey of existing heater designs. The empirical formulas usually yield satisfactory results for folded-heater designs, although discrepancies of about 5 to 10 per cent may result when the empirical formulas are applied to designs of single-helical or double-helical heaters. The calculations lack high precision because only heater wire dimensions are taken into consideration; other smaller but significant effects are neglected. Among the minor factors not specifically taken into account -- any of which could give rise to a 2 or 3 per cent discrepancy — are:

1. Heater type (folded, single helical, double-helical) and temperature distribution (The calculations are most reliable for folded heaters.)
2. Thickness of insulation coating
3. The percentage of the heater length exposed outside the cathode at the legs and apices
4. The percentage of heater area that can "view" the cathode to permit direct radiation coupling
5. The effects of processing (severe processing can increase heater current 10 to 15 per cent)
6. Cathode materials and mount structures
7. Wire type, i. e., whether Dowmo or tungsten—the resistance is practically the same at high temperature
8. The "whiteness" (i. e., total emissivity) of different lots of heater coatings

Heater Wire Dimensions

The actual calculation of heater wire dimensions by means of the nomographs (Figs. 1, 2, and 3) is illustrated by the example in Fig. 1. The first step is to select a normal operating temperature (the effects of temperature on heater reliability are discussed in a later section). This temperature will usually be in the range 1450 to 1600 K, but it is generally desirable to choose a temperature as low as possible (below 1575 K) to assure good reliability. A straight line is then drawn through the scales corresponding to the desired heater current and the selected operating temperature. The intercepts of this line with the other scales give the appropriate wire size and the heater length in millimeters per volt. From the value for heater length, the lighted length can be determined. It is then necessary to determine by calculation or experiment whether the length of heater wire when folded or coiled and coated with an appropriate amount of insulation (as discussed subsequently) will fit into the cathode. If the heater is too large, it will be necessary to redesign it to be smaller and to have a higher operating temperature. If, on the other hand, the heater would fit too loosely in the cathode, it is generally desirable to select a "cooler" heater with greater reliability.

Because heater wire diameter cannot be measured directly with accuracy, specifications for heater wire size are based on the wire weight per 200 millimeters. Fig. 2 is a nomograph which permits rapid calculation of this quantity from the diameter of a wire of known density.

Fig. 3 is a nomograph useful in the design of coiled heaters. It permits the rapid determination of the proper number of turns per inch once the mandrel diameter and the heater wire dimensions have been selected.

Although is is possible to calculate all necessary heater parameters from theoretical considerations, the

empirical nomographic method is generally recommended for routine heater design. The theoretical calculations are laborious and, unless accurate estimates of end losses and other correction terms are available, may prove no more reliable than the rapid empirical method. The basic features of a simplified sample calculation are, however, of interest in that they permit a physical picture to be drawn of the principles underlying heater operation and design.

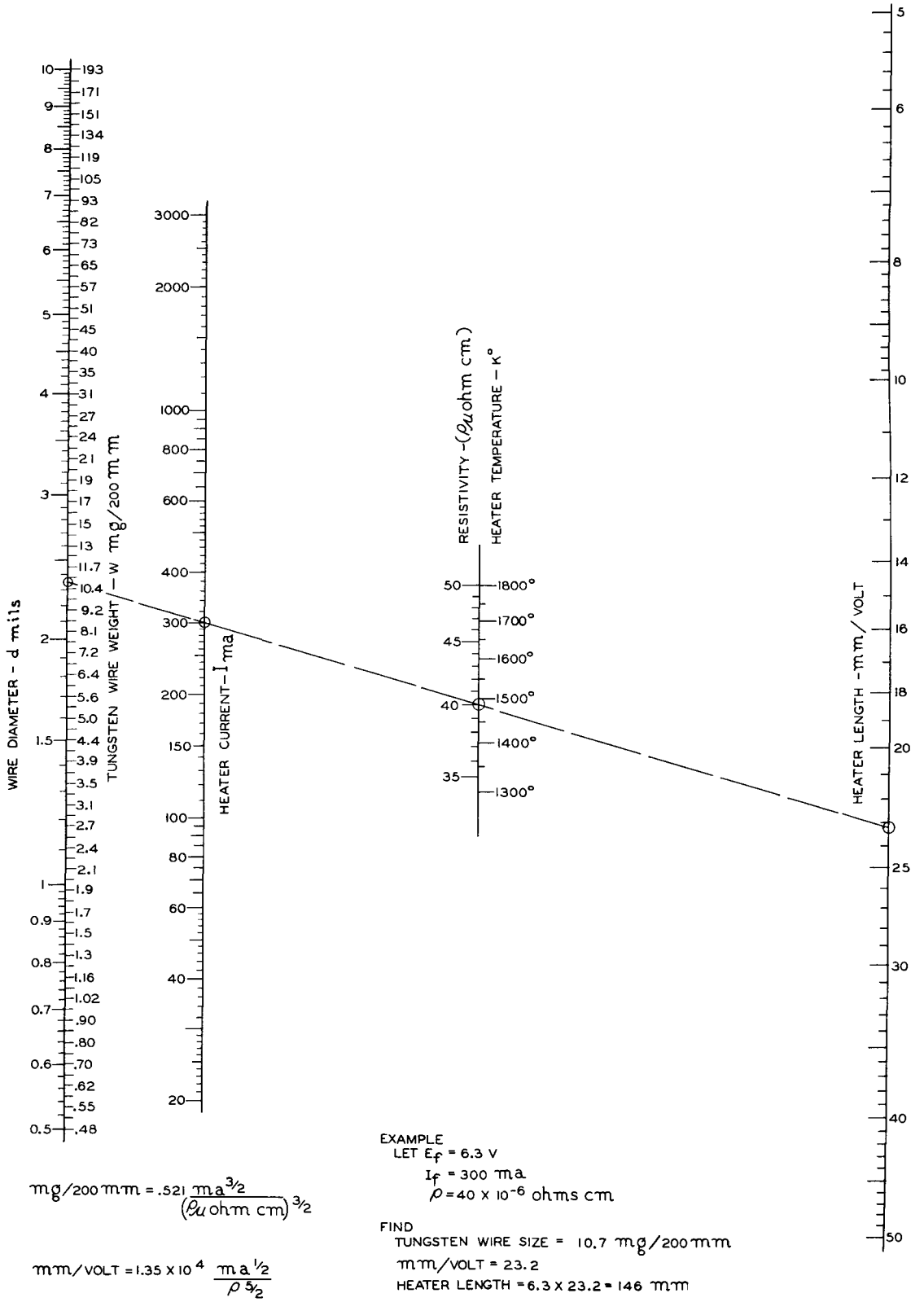

WIRE DIAMETER - d mils
TUNGSTEN WIRE WEIGHT - W mg/200 mm
HEATER CURRENT - I ma
RESISTIVITY - (ρu ohm cm)
HEATER TEMPERATURE - K°
HEATER LENGTH - mm/VOLT
EXAMPLE
LET Ef = 6.3 V
If = 300 ma
ρ = 40 x 10⁻⁶ ohms cm
FIND
TUNGSTEN WIRE SIZE = 10.7 mg/200 mm
mm/VOLT = 23.2
HEATER LENGTH = 6.3 x 23.2 = 146 mm
mg/200 mm = .521 ma^3/2 / (ρu ohm cm)^3/2
mm/VOLT = 1.35 x 10^4 ma^1/2 / ρ^5/2

Figure 1. Empirical Tungsten Heater Design in Terms of Current, Voltage, Wire Weight, and Resistivity

The important parameters of the heater materials

(at their normal operating temperature of 1500 K) currently used throughout the electron-tube industry are summarized in Fig. 4.

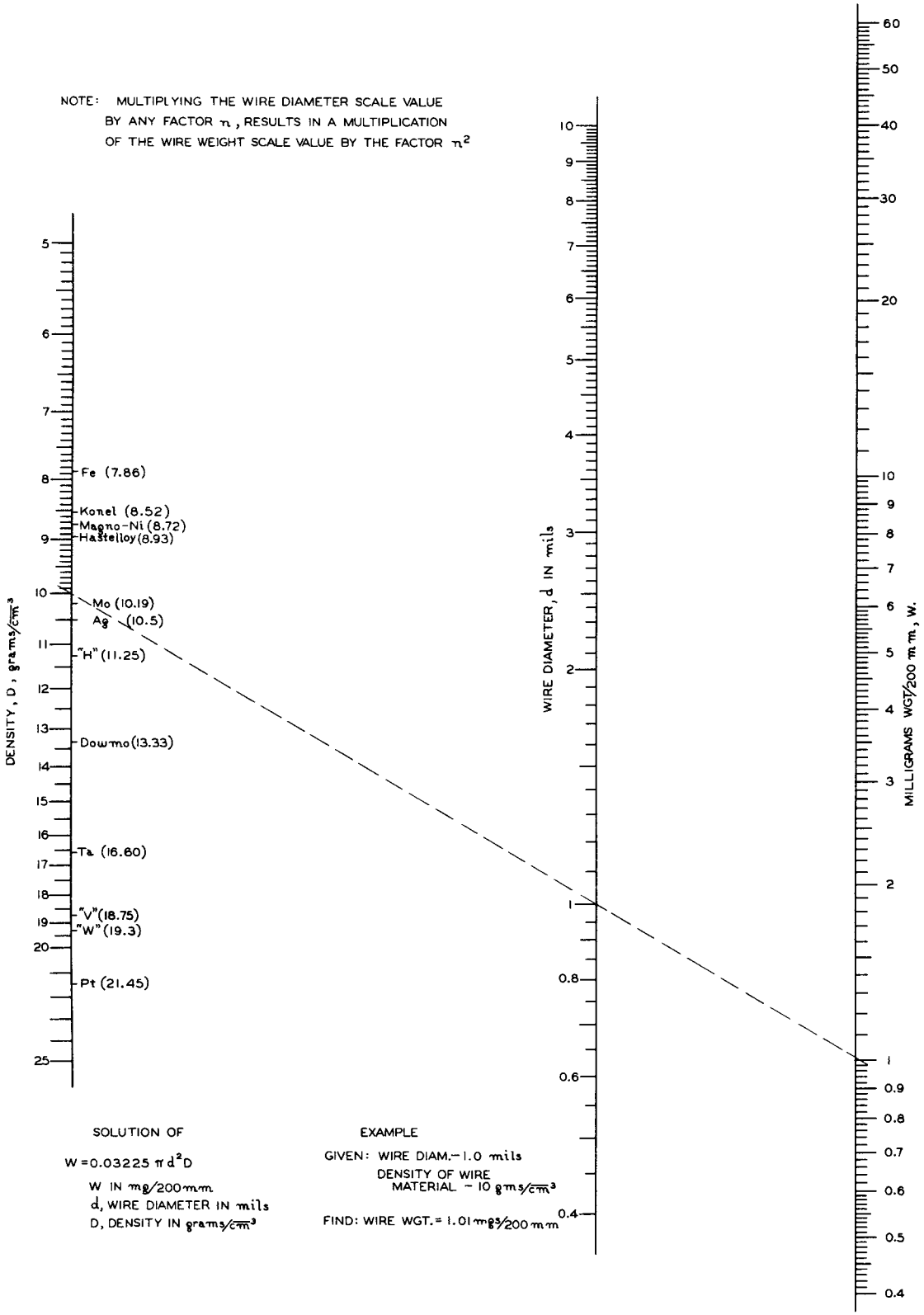

NOTE: MULTIPLYING THE WIRE DIAMETER SCALE VALUE BY ANY FACTOR n, RESULTS IN A MULTIPLICATION OF THE WIRE WEIGHT SCALE VALUE BY THE FACTOR n²
DENSITY, D, grams/cm³
Fe (7.86)
Konel (8.52)
Magno-Ni (8.72)
Hastelloy (8.93)
Mo (10.19)
Ag (10.5)
"H" (11.25)
Dow-mo (13.33)
Ta (16.60)
"V" (18.75)
"W" (19.3)
Pt (21.45)
WIRE DIAMETER, d IN mils
MILLIGRAMS WGT/200 m m, W.
SOLUTION OF
W = 0.03225 π d²D
W IN mg/200 mm
d, WIRE DIAMETER IN mils
D, DENSITY IN grams/cm³
EXAMPLE
GIVEN: WIRE DIAM. = 1.0 mils
DENSITY OF WIRE MATERIAL = 10 gms/cm³
FIND: WIRE WGT. = 1.01 mgs/200 mm

Figure 2. Round Wire: Weight, Diameter, and Density

Recent studies by RCA engineers indicate that the mechanism of heat transfer to the cathode is a two-stage process (Fig. 4). First, the thermal energy of the electrically-heated tungsten wire is transferred to the insulation, with which it is in intimate contact, through conduction. Due to the small dimensions and relatively high conductivity of the insulation, this proc-

ess is highly efficient and, therefore, temperature gradients between the surface of the heater coating and the adjacent heater wire will generally be less than one

by radiation across the vacuum separating the heater and cathode. As indicated schematically in Fig. 4, conduction effects are not important because the area of

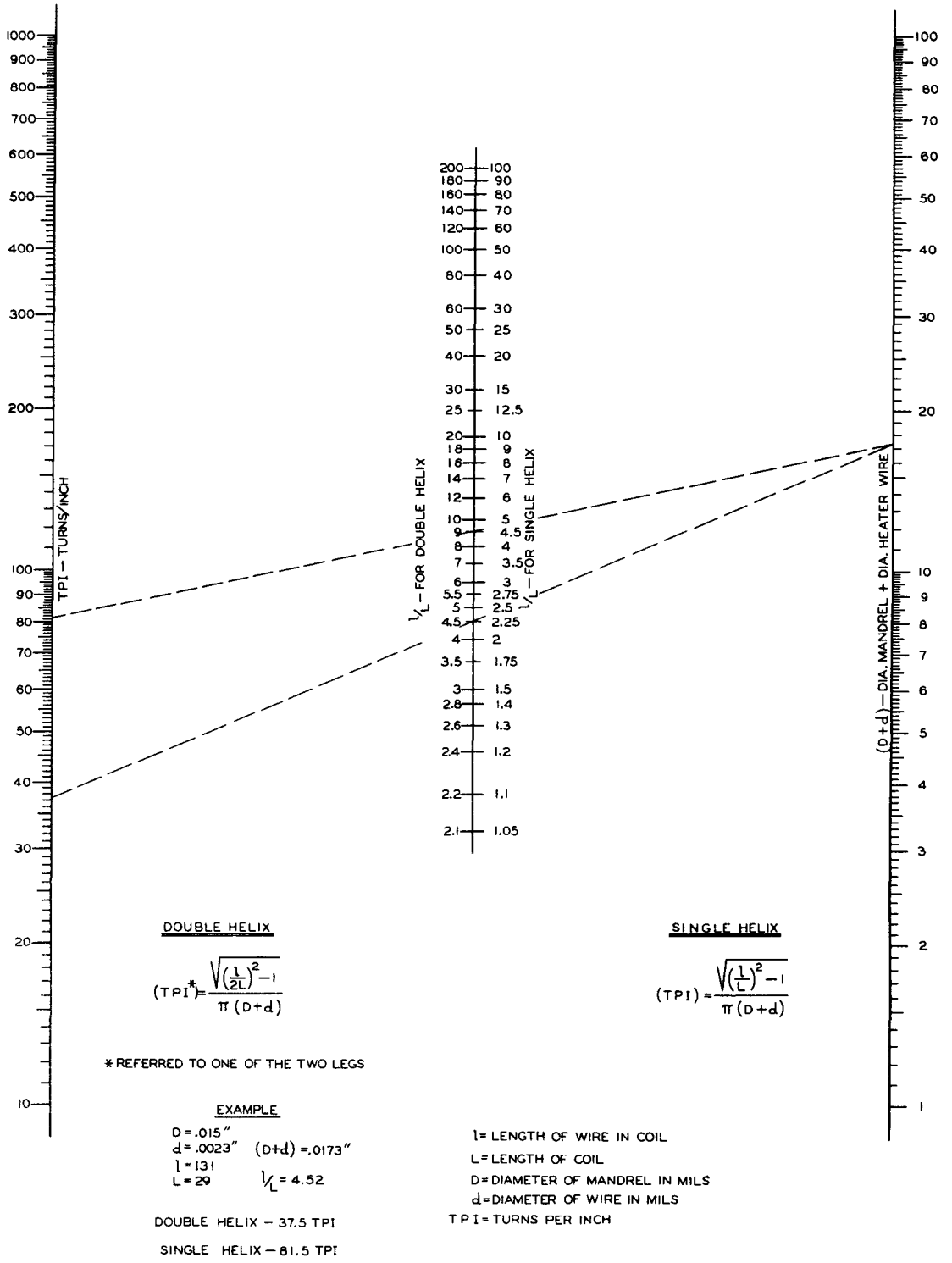
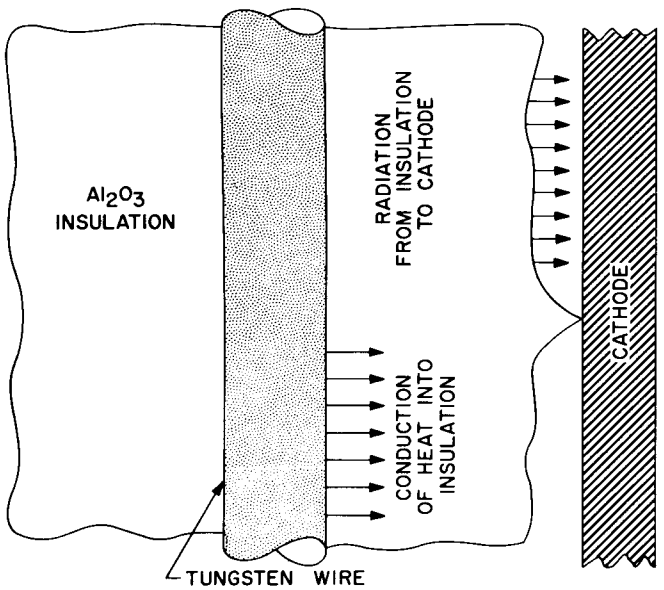

TPI — TURNS/INCH
l/L — FOR DOUBLE HELIX
l/L — FOR SINGLE HELIX
(D+d) — DIA. MANDREL + DIA. HEATER WIRE
DOUBLE HELIX
$(TPI)^{*}=\frac{\sqrt{\left(\frac{l}{2L}\right)^2-1}}{\pi(D+d)}$
SINGLE HELIX
$(TPI)=\frac{\sqrt{\left(\frac{l}{L}\right)^2-1}}{\pi(D+d)}$
*REFERRED TO ONE OF THE TWO LEGS
EXAMPLE
D = .015"
d = .0023" (D+d) = .0173"
l = 131
L = 29 l/L = 4.52
DOUBLE HELIX — 37.5 TPI
SINGLE HELIX — 81.5 TPI
l = LENGTH OF WIRE IN COIL
L = LENGTH OF COIL
D = DIAMETER OF MANDREL IN MILS
d = DIAMETER OF WIRE IN MILS
TPI = TURNS PER INCH

Figure 3. Single- and Double-Helical Heaters: winding characteristics

or two degrees Kelvin. Second, the thermal energy arriving at the surface of the insulation is transferred

contact between the granular heater coating surface and the cathode is small. Radiation coupling of the heater

insulation and the cathode surface is by no means as "efficient" as conduction coupling, and an average temperature gradient of 400 to 500 K exists between these two surfaces at equilibrium. Such large temperature differentials, although to a certain extent characteristic of radiation coupling, are mainly due to the low total emissivity coefficient of the insulation coating — only 18 per cent of that associated with an ideal black body.



TYPICAL PHYSICAL CONSTANTS OF HEATER MATERIALS AT 1500 K

Al_2O_3 INSULATION

TOTAL EMISSIVITY	= 0.17–0.20
SPECTRAL EMISSIVITY	= 0.13
CONDUCTIVITY (WATT·CM^{-1}·DEG K^{-1})	= 0.01

TUNGSTEN WIRE

TOTAL EMISSIVITY	= 0.19
SPECTRAL EMISSIVITY	= 0.45

Figure 4. Schematic Representation of the Heat Transfer Processes Between Heater and Cathode, and the Appropriate Heat-Transfer Coefficients

To calculate the heat balance at equilibrium conditions, it is necessary to equate the radiation losses from the surface of the insulation coating with the thermal energy generated electrically in the heater wire. To an excellent degree of approximation, the losses from the heater may be assumed to result entirely from radiation so that the Stefan-Boltzman relation will apply:

$$P = E_f \cdot I_f = \sigma \epsilon S (T_h^4 - T_c^4) \quad (1)$$

where P = power supplied to the heater (watts)
E_f = heater voltage (volts)
I_f = heater current (amperes)
σ = universal constant ($5.673 \cdot 10^{-5}$ erg·cm^{-2}·deg^{-4}·sec^{-1})
T_h = heater temperature (deg K)
T_c = cathode temperature (deg K)
S = effective surface area of heater insulation (cm^2)
ϵ = effective emissivity of heater-cathode system

Since the cathode, in a similar manner, will lose by radiation most of the heat supplied to it, it is readily proved that the cathode temperature term T_c^4 is essentially proportional to the heater temperature term T_h^4 and Eq. (1) may, therefore, be simplified to contain only heater temperature terms:

$$P = E_f \cdot I_f = K S T_h^4 \quad (1a)$$

where K is a radiation parameter which is nearly constant over a wide temperature range. The accuracy of the radiation hypothesis resulting in Eq. (1a) may be tested experimentally as shown in Table I for a 6J4WA tube. The slight discrepancies between actual power dissipations and calculated power dissipations are mainly due to the neglected effects of end losses and changes of radiation parameter K with temperature.

Table I

Measured and Calculated Power Dissipation from a 6J4WA Assuming Heat Transfer by Radiation

Heater volts	Power Dissipation — Watts	
	Actual	Calculated
0.00	0.00	0.00
0.99	0.16	0.13
2.00	0.46	0.43
6.30	2.68	2.68
12.00	7.30	7.24

Because the effective radiating surface area S in Eq. (1a) depends on the exposed area of insulation rather than on the heater wire area, it is not possible, in general, to make refined calculations of heater characteristics in terms of wire dimensions alone. Fortunately, only a limited range of all possible permutations of wire diameters, coating thicknesses, cathode and heater temperatures, cathode sizes, etc., will be found in receiving tubes of conventional design. As a result, an approximation, which can be shown to be reasonably consistent with conventional folded-heater design, is that the effective radiating area is proportional to the surface area of the heater wire. It may be noted that the success of the empirical nomographic method of heater design depends on the approximate validity of a relationship of this type. With such an assumption, a suitable radiation-type estimate of the power dissipation from a folded heater is given by:

$$P = E_f I_f = 1.1 \cdot 10^{-15} DLT^4 \quad (1b)$$

where D is the diameter of the wire in thousandths of an inch, L is the effective lighted length of the wire in millimeters, and T is the temperature in degrees Kelvin.

A second, exact, relationship, entering into heater design, which specifies the input electrical power, is given by the relationship between wire size and the average resistivity of the heater wire. Because the average resistivity of the heater wire depends on the heater temperature, this equation introduces heater temperature indirectly into design considerations.

$$\rho = \left(\frac{E_f}{I_f}\right)\left(\frac{A}{L}\right) \quad (2)$$

or, for tungsten wire:

$$\rho = 26.1\left(\frac{E_f}{I_f}\right)\left(\frac{w}{L}\right) \times 10^{-6} \quad (2a)$$

where ρ = average resistivity of the heater wire (ohm-cm) from which the temperature is determined according to Table II
A = cross-sectional area of heater wire (cm^2)
E_f = heater voltage (volts)
I_f = heater current (amperes)
w = weight of tungsten wire (milligrams/200 mm)
L = effective lighted length of wire (mm), i.e., the length of heater inside the cathode plus half the remaining lighted length

Table II

Resistivity of Tungsten at Different Temperatures

Temperature — deg K	Resistivity — ohm-cm
1200	30.98×10^{-6}
1300	34.08
1400	37.19
1500	40.36
1600	43.55
1700	46.78
1800	50.05
1900	53.35
2000	56.67
2100	60.06
2200	63.48
2300	66.91

By simultaneous solution of Eq. (1b) and (2a) with the aid of Table II, it is possible to calculate heater designs on a partially theoretical basis. Table III shows a comparison of heater currents calculated (by both theoretical and empirical methods) for four 6.3-volt heaters selected at random. Although the values calculated by the theoretical method compare favorably with those calculated by the empirical method, there is apparently in both cases a systematic deviation at the high and low end of the heater-current scale.

Table III

Comparison of Different Methods of Heater Design Calculation

Heater	Heater volts	Heater Currents-amperes		
		Nomograph	Theory	Rated
FH870A	6.3	0.204	0.196	0.200
FH795B	6.3	0.306	0.299	0.300
FH761B	6.3	0.604	0.600	0.600
FH542B	6.3	1.21	1.28	1.20

The importance of the theoretical methods of heater design is mainly that they permit a clearer picture of heater operation to be drawn. The thermal energy supplied by the heater to the cathode is largely in the form of radiation. Exactly the same amount of energy can be supplied either by heater wires with a high operating temperature and a small wire surface area (i.e., heaters of small size) or by heaters of low temperature and large wire surface area (i.e., heaters of large size). It may be further seen that any factor which would alter the effective emissivity coefficient K of the heater coating would have a substantial effect on the power dissipation.

Heater Design Temperature

Average heater temperature may be calculated without any special assumptions from Eq. (2a) and Table II on the basis of heater-wire dimensions and the current corresponding to a given heater voltage. Temperatures calculated in this manner are usually very close to those obtained by "corrected" optical temperature measurements. The average operating temperature is an extremely important parameter of heater design and should be calculated at the first stages of attacking any heater problem.

The equations for the design of heaters show that there are an infinite number of possible combinations of wire surface area and wire temperature which will result in a desired power dissipation. It is necessary, therefore, to specify heater temperature in order to fix the heater design. Since the size of a heater increases rapidly as the operating temperature is lowered, there will be some limiting design temperature below which it will not be possible to fit a proposed heater into the desired cathode. The heater temperatures used in miniature tubes are generally in the range of 1450 to 1600 K with most miniature tube heaters designed for temperatures below 1575 K.

The over-all reliability of heaters decreases very rapidly with increasing design temperature and, therefore, it is usually desirable to make the heater design temperature as low as possible. With folded heaters having wire sizes in the range of 2.0 mils, the temperature effect can be especially critical because the mechanical strength of such fine wires is low — especially if the apices have been split during the folding operation or if extensive recrystallization of the wire has occurred. A typical experimental estimate of the effect of operating temperature on heater reliability is that an increase in heater operating temperature of 50 K will decrease reliability by a factor of three or four. Experience has repeatedly shown that heaters operating at temperatures below 1550 K are usually highly reliable and any difficulties experienced in normal life are usually due to mechanical causes. When design temperatures of 1600 K are exceeded, the effects of temperature become more pronounced and occasional life-test or field problems may be encountered. Tubes designed with heaters operating near 1700 K should be viewed with considerable caution; unless the heater-wire diameter is large, careful factory processing may be required to ensure satisfactory life.

Two possible exceptions to the general desirability of using "cool" heaters, however, may be noted. First, it has been reported that the slump in transconductance for tubes operating between 100 per cent and 90 per

cent of rated heater voltage can be slightly reduced by using a "hot" heater. Second, it is generally more difficult to age out heater-cathode leakage and hum from the large "cool" heaters, as will be discussed later.

Insulation Thickness

The technical difficulties of applying a uniform, thin layer of insulation to fine wire are considerable; most current specifications allow for a variation of approximately ±7.5 per cent from the desired average coating thickness. By utilizing phototube methods of diameter control or by additional inspection, this tolerance can be reduced to better than ±5 per cent. In order that the effects of coating thickness tolerances can be eliminated from discussions of the insulation necessary on a heater, however, the insulation thickness will be defined here, not as the average, but as the minimum specification limit for the distance between the heater wire surface and the outside of the coating.

The amount of insulation required by a heater depends mainly on its heater-cathode voltage ratings. A conservative and reliable rule of thumb requires one mil of insulation for every 75 volts of dc voltage (or peak ac voltage) or a two-mil minimum thickness (that is, t = 2.0 mils or $E_{HK}/75$ mils whichever is greater). Thicknesses corresponding to 145 volts per mil have been used, but voltage gradients as high as these leave little margin for error in heater processing. Fig. 5 shows the results of a survey taken from life test records of commercial tubes over several months vs. calculated heater temperature. It will be noted that the apparent conservative minimum amount of insulation thickness corresponds closely with the 75-volts-per-mil rule.

Use of somewhat more than the minimum amount of insulation is generally beneficial and, therefore, most tube types are designed with more insulation than necessary for the heater-cathode ratings. The additional coating thickness not only provides an extra margin of safety against insulation breakdowns during life, but also minimizes the possibilities of apex-to-apex shorts or apex-to-cathode shorts with folded heaters. It should be noted, however, that increased coating thickness will not usually improve heater-cathode leakage (as measured at a bias voltage of 100 volts), and in many instances may cause additional heater failures on intermittent life if the strength of the coating exceeds that of the wire so that the wire can pull apart before the coating cracks.

A final, and often decisive, factor in the selection of a suitable insulation thickness is the heater warm-up time desired. As discussed subsequently, increased coating thickness results in increased heater mass and, as a result, in increased warm-up time.

Apex Chip Size

In the manufacture of folded heaters, a wire continuously coated with insulation is wound over sharp blades. During this process, the coating is knocked off at the apices where the folds have been made so that areas of bare wire are exposed. The length of bare wire exposed is defined as the apex chip size. Because the exposed areas of heater wire are not normally recoated (except for premium tubes), it is desirable that the apex chip size be kept as small as possible in order to prevent shorts between heater and cathode. In addition, because the uncoated apicies of the heater run substantially hotter than the insulated areas, it is desirable to minimize the chip size. No corresponding problems exist, of course, with coiled heaters since the insulation is applied after the coils have been formed.

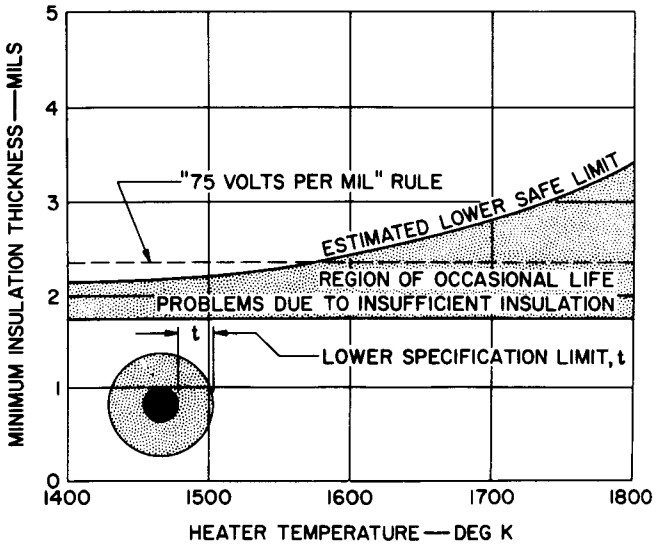


Figure 5. Results of Life-Test Survey Showing Apparent "Safe" Lower Specification Limit for Insulation Thickness for a Heater-Cathode Voltage of 180 Volts DC During Life

Apex chip size depends on many factors including coating hardness, coated-wire-to-bare-wire diameter ratio, and machine maintenance. The chip size is usually about 1.5 times the coating diameter.

While the bare apices of folded heaters may be recoated in an additional manufacturing operation, most commercial tubes use heaters, as wound, with bare apices. In order to prevent the occurrence of apex-to-apex shorts, it is necessary to design these heaters in such a manner that two bare apices cannot fall opposite each other. The method used is to make each heater leg progressively shorter (i.e., a staggered heater) or to lower successive heater legs a fixed amount while the average leg length is kept constant (i.e., a sloped heater).

Heater Warm-Up Time

The use of tubes in series-string circuits in television and radio receivers has necessitated controls on the warm-up time of heaters. If, for example, all the heaters in a string do not warm up uniformly, excessive heater voltage surges will be developed across the faster-heating tubes because the electrical resistance of the other heaters is low until they reach operating temperature.

The circuit which has been standardized throughout

the electron-tube industry for measuring the warm-up time of tubes is shown in Fig. 6.

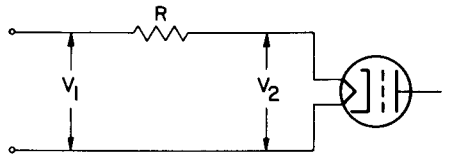


Figure 6. Basic Circuit Used in Measurement of Heater Warm-Up Time: V_1 = four times rated heater voltage (to nearest volt) applied to circuit; R = three times hot resistance (E_f/I_f) to nearest ohm; V_2 = voltage across heater

This circuit includes a resistance three times the hot resistance of the heater (to the nearest ohm) in series with the tube heater. To this circuit is applied an ac potential equal (to the nearest full volt) to four times the rated heater voltage of the tube. Under these conditions, warm-up time is defined as the time in seconds required for 80 per cent of the rated voltage to be developed across the tube after the power has been applied. This circuit roughly duplicates the actual conditions of series-string operation. Present specifications call for a warm-up time of 11 seconds for series-string types.

The warm-up time of a tube in such a circuit may be calculated with considerable accuracy from a knowledge of heater dimensions only[2]. Warm-up time is approximated by the equation:

$$t = 4.18\int_{300}^{T} \frac{W_i\,[C_p\,(T)_i] + W_w\,[C_p\,(T)_w]}{\left\{\dfrac{E_f}{\dfrac{L\,[\rho(T)_w]}{A} + \dfrac{3E_f}{I_f}}\right\}^2 \left(\dfrac{L\,[\rho\,(T)_w]}{A}\right) - KT^4}\,dT$$

where:

T = temperature (deg K)

W_i = weight of insulation coating (grams)

W_w = weight of wire (grams)

$C_p(T)_i$ = specific heat of insulation coating (temperature-dependent) (calories/gram)

$C_p(T)_w$ = specific heat of wire (temperature-dependent) (calories/gram)

$\rho\,(T)_w$ = specific resistivity of wire (temperature-dependent) ohm-cm

L = length of wire (cm)

A = cross-sectional area of heater wire (cm^2)

E_f = applied voltage (volts)

Fig. 7 shows typical calculated and measured warm-up time characteristics for a 6AU6 tube. Although the agreement might be sufficiently accurate for design purposes, the calculation is lengthy. It is preferable in practice to adjust warm-up times from experimental data.

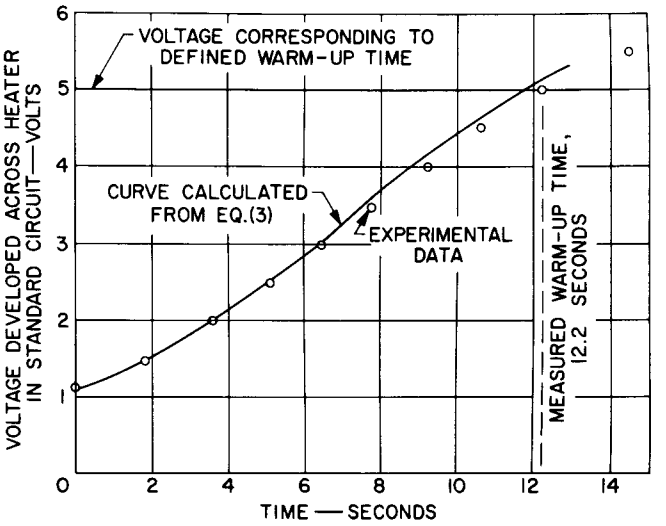


Figure 7. Calculated and Measured Warm-Up Time Characteristics of a 6AU6 Type Miniature in the Standard Heater Warm-Up Circuit

Analysis of Eq. (3) shows that the warm-up time for a heater of a given current rating will be increased by three factors: (1) increased heater coating weight or thickness (more mass to heat up); (2) low design temperature ("cool" heaters are larger than "hot" heaters and require more time to heat up); (3) below-normal heater current (less power is supplied to low-heater-current tubes and, therefore, warm-up time is greater).

Approximate solutions of Eq. (3) show that, basically, the total warm-up time may be viewed as consisting of a short period of approximately one to three seconds required for warming up the wire mass and a longer time, directly proportional to the weight of heater coating, required for warming up the insulation mass.

Fig. 8 illustrates this point by a comparison of the calculated and measured warm-up times for a series of 6AU6 tubes with different coating weights.

If data for the warm-up time of a prototype heater of known wire weight and insulation coating weight are available, then the warm-up time of the same heater when coated with a different coating weight may be estimated from a simple ratio.[2] Since insulation coatings are specified in terms of thicknesses instead of by weight, it may be more convenient to use coated diameters instead of coating weights as a basis for calculation. Then, for folded heaters having an insulation coating density of 1.8 grams per cubic centimeter, which is typical of some drag-coated heaters, the calculated effect of coated diameter on warm-up time becomes:

$$\frac{t_1}{t_2} \cong \frac{D_1^2 + 0.4d^2}{D_2^2 + 0.4d^2} \qquad (4)$$

where: d = wire diameter (mils)
D = coated diameter (mils)

t_1 = warm-up time corresponding to coated diameter D_1

t_2 = warm-up time corresponding to coated diameter D_2

Eq. (4) indicates that warm-up time of heaters is mainly a function of thickness of insulation coating and that in order to control the warm-up time accurately it is necessary to control the amount of insulation coating accurately. For this reason, phototube and other controls have been adopted by the factory to control the diameter of drag-coated wire to better than ±5 per cent on series-string tube types.

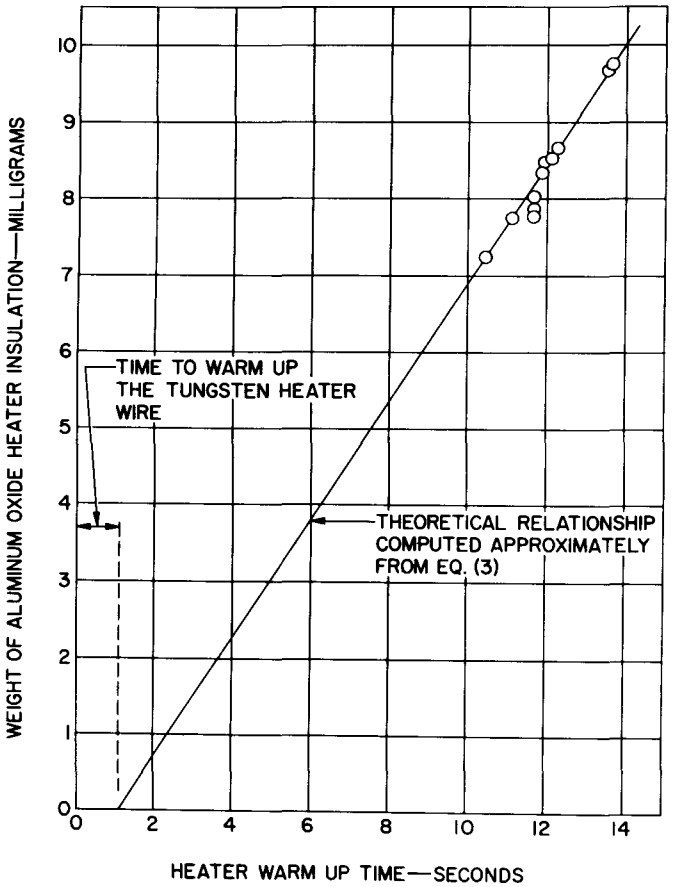


Figure 8. Calculated and Measured Dependence of Warm-Up Time on the Weight of Heater Insulation for a Series of Type 6AU6 Miniature Tubes Drawing Exactly 0.300 Ampere Heater Current

HEATER TYPES

Three basic types of heaters are currently used in receiving tubes: (1) folded heaters; (2) folded single-helical heaters; (3) double-helical heaters. Previous design considerations apply to all of these types. The three types are illustrated in Fig. 9.

Folded Heaters

Drag-coated folded heaters have received widespread use during recent years. Because of the simplicity of manufacture, the cost of folded heaters is substantially less than that of other types. Basically, these heaters are made by taking an appropriate length of insulated heater wire from a spool and folding it. Consequently, relatively little handling is required and the cost is low. On the other hand, the uniformity in electrical insulation quality of folded heaters tends to be poorer than in the case of the coiled heaters. This effect appears to be mainly due to the fact that folded heaters do not normally receive a high-temperature firing in hydrogen during manufacture. As a result, the tungsten wire may be oxidized to an undesirable extent, and the coating may contain contaminants which have not been vaporized off during a high-temperature treatment. Although the reliability of folded heaters is reputed to be somewhat lower than that of the other two types, this difference, in practice, need not be large.

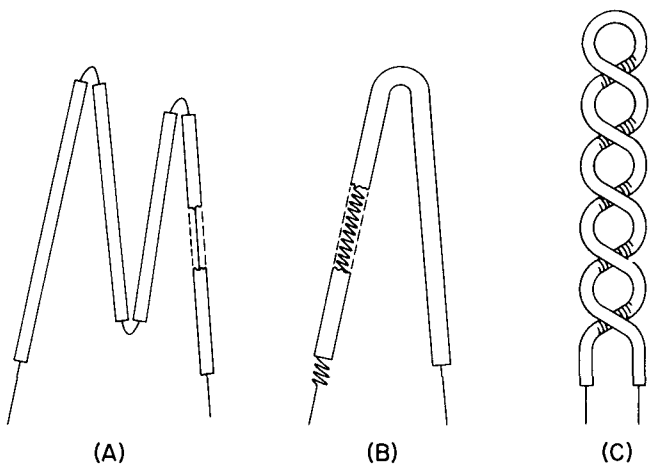


Figure 9. Heater Types Used in Electron Tubes: (A) folded heater; (B) folded single-helical (hairpin) heater; (C) double-helical heater

Folded heaters may be used in any size or shape of cathode, either round or flat, and are at present the most economically satisfactory choice for most tubes. Because of the flexibility in choosing the number and length of the folds, it is often possible to achieve more uniform temperatures along the cathode than can be obtained with coiled heaters. The less desirable features are: (1) relatively inefficient packing of the heater wire inside the cathode necessitating high heater design temperatures in some cases, (2) a certain degree of magnetic coupling between heater and the electron stream which can cause some hum (usually negligible for all but low-noise tubes), (3) higher mounting shrinkage, (4) areas of bare wire at the apices, (5) sharp bends at apices may split the wire, and, hence, be the site of "open" heater failures. This effect is made more severe by the circumstance that the bare apices run hotter than the coated portions of the heater wire because of the smaller surface area from which power can be radiated.

Double-Helical Heaters

Double-helical heaters were at one time used exclusively in the manufacture of receiving tubes, but have gradually been replaced by folded heaters in most tube

types due to the higher cost of double-helical heaters. Although this type of heater has a few disadvantages, experience indicates that on the whole, double-helical heaters approach the optimum in heater design. In tube types where the lowest possible hum levels must be achieved, it is usually desirable to use this type of heater.

The higher cost of double-helical heaters is due to the fact that each heater must be handled individually several times during manufacture. Individual handling is required after the wire coiling operation, after the insulation is applied by spraying, and after the insulation is sintered onto the wire by firing in wet hydrogen at approximately 1700 C.

The double-helical heater is potentially superior to folded heaters because of the absence of sharp bends and bare apices (since the insulation is applied after the coil is formed), and since magnetic coupling between the heater and the electron stream, which could cause some magnetic hum, is almost perfectly neutralized. Furthermore, the double-helical type is reputed to show generally lower levels of heater-cathode leakage and of hum of the type produced by leakage. The latter result is probably connected with the high-temperature treatment in hydrogen which cleans the heater, and possibly also to the circumstance that there are fewer points of contact between the insulation coating and the cathode than in the case of folded heaters.

It is often especially advantageous to consider the double-helical heater for use in applications where folded heaters with cataphoretically coated apices would be required. Among the disadvantages of double-helical heaters at present are: (1) higher cost; (2) a round shape which usually restricts use to round cathodes; (3) technical difficulties in winding small coils, which restricts the use to cathodes 30 mils in diameter or larger; and (4) chipping of the coating on the heater legs near the bottom of the cathode on 1 to 4 per cent of the tubes manufactured.

Folded Single-Helical (Hairpin) Heaters

Folded single-helical heaters possess many of the advantages of the double-helical heater. Until recently this type of heater was not widely used in this country, although foreign manufacturers have made wide use of them. The main drawbacks of this type are a somewhat higher cost than folded heaters together with somewhat increased welding and insertion problems.

In the manufacture of folded single-helical heaters, a tungsten wire is continuously wound upon a molybdenum mandrel. The coils may be wound either continuously or with suitably spaced "skips" for the heater legs and apices. The skip design, although more costly, is preferable since it enables the coiled sections of the heater to be concentrated within the cathode sleeve and is found to minimize twisting of the heater during life. Without such a feature, the heater current of finished tubes can be variable. After the winding operation, the coil-covered mandrels are cut into appropriate lengths, folded, and coated cataphoretically with aluminum oxide insulation. The aluminum oxide insulation is then fired in wet hydrogen at 1700 C; finally, the mandrel is dissolved out in acid baths. Occasional factory problems have been encountered when the mandrel was not completely removed, but such instances are rare.

The leakage, hum, and life characteristics of folded single-helical heaters are generally superior to those of the less costly folded heaters. Somewhat better compensation of magnetic field is provided by the folded single-helical heater design. (This compensation is, however, not as perfect as with the double-helical heater designs.) In addition, the high-temperature treatment in hydrogen vaporizes away many contaminants and controls the amount of oxide formed on the tungsten wire. Finally, the absence of bare apices and sharp bends minimizes emission from the tungsten wire and is conducive to long-life operation. It is anticipated that this type of heater will find increased application, especially where folded heaters with cataphoretically coated apices have been previously used.

HEATER MATERIALS

Heater Wire Materials

Heaters normally operate at temperatures in the range of 1450 to 1600 K; during processing, temperatures of 2000 to 2300 K are often reached for short periods. There are only a very few metals capable of withstanding such severe conditions. Of these, technical and economic considerations presently eliminate all but tungsten and molybdenum. Rhenium heaters have been reported as feasible in the literature,[3] but this exceedingly rare and costly element is not now commercially available, and certain technical difficulties associated with the commercial drawing of fine wires have not yet been overcome.

The hot strength of pure molybdenum is so low under typical heater application conditions that molybdenum can only be used when alloyed with substantial amounts of tungsten. Years of experience have established two basic types of wire as satisfactory for heaters: (1) tungsten; and (2) Dowmo ("J" wire), an alloy containing 50 per cent molybdenum and 50 per cent tungsten. While the resistivity of these materials is not the same at room temperature, there is relatively little difference at normal heater operating temperatures so that Dowmo and tungsten heaters of the same dimensions are interchangeable. Dowmo, however, is less dense than Tungsten so the wire weight of tungsten will be 1.46 times that of a Dowmo wire of the same diameter. Although Dowmo heaters have certain manufacturing advantages as compared with tungsten, heaters made from tungsten invariably show better life performance due to a superior hot strength. Consequently, Dowmo has gradually been replaced by tungsten for most receiving-tube applications.

Tungsten as a Heater Material

Tungsten has the highest melting point (3410 C) and when in the fibrous state, the highest tensile strength (approximately 500,000 pounds per square inch) of the known metals. Other features of tungsten, however, would not be desirable in an ideal heater wire material:

(1) embrittlement takes place after prolonged high-temperature operation; (2) the coefficient of expansion (about 5 x 10^{-6} per deg C) does not match that of the alumina insulation coating (about 8 x 10^{-6} per deg C); (3) the electron emission at normal heater temperatures is great enough to contribute noise and hum; and (4) volatile oxides, which can contaminate and decrease the insulation value of the heater coating, are easily formed.

The high strength shown by tungsten in the fibrous state largely disappears when it is exposed to high temperatures for prolonged periods. The reduction in strength results from recrystallization and the growth of individual tungsten crystals. The bonding between large crystals is extremely weak and, therefore, when recrystallization has progressed to the extent that single crystals cover entire cross sections of wire, the wire becomes too brittle to handle without breaking. On intermittent operation, the nonuniformity of heating and the mismatch of thermal expansion coefficients of the wire and the insulation produce mechanical stresses as the wire heats and cools. Open heater failures, especially with the smaller wire sizes, can result if the wire strength deteriorates; the amount of deterioration depends on the amount of recrystallization which has occurred.

The rate of recrystallization of tungsten wire depends on time, temperature, previous processing history, and the presence of trace amounts of additives (e.g., alkali silicates and aluminates) which control crystal growth. In addition, there are certain contaminants (e.g., nickel, which may be deposited on the heater during life) that severely weaken the tungsten if present in even minute amounts. Due to the important effects of these trace impurities and the processing history, the manufacture of tungsten wire is difficult to control and large differences in wire quality may, therefore, exist among spools of wire supplied by the same manufacturer. Wire quality, with respect to recrystallization temperature may be roughly evaluated by "flashing" small samples to a high temperature for a short time and then measuring tensile strength or examining the crystal structure of polished sections. Fig. 10 shows photomicrographs of the crystal structure of tungsten wire in its original fibrous state and after flashing to different temperatures; the tensile strength measurements of samples corresponding to those shown in Fig. 10, are shown in Fig. 11 (A). The results are typical of the loss of strength due to recrystallization.

Fig. 11 (B) shows the percentage of original tensile strength of samples of tungsten wires supplied by different manufacturers, after flashing to 62 per cent of fusion current for 60 seconds. The large range in strengths of the different tungsten wires indicates clearly that not all tungsten wires are the same, and that certain types show superior performance. This statement is supported by life data.

Due to the ease with which tungsten oxidizes and with which the oxide can diffuse into and contaminate the insulating coating, it is necessary to take care, during the manufacture and processing of heaters, to avoid oxidation of the wire. As discussed subsequently, these oxides can cause serious deterioration of the "insulation value" of the coating. The following oxides have been reported in the literature[4] WO, WO_2, WO_3, W_2O_3, W_2O_5, W_3O_8, W_4O_3, W_4O_{11}, W_5O_8, W_5O_{11}, W_5O_{14}, indicating that tungsten can take up oxygen in almost any proportion. High temperatures are not necessary for oxidation of heaters; tungsten wire will slowly oxidize to WO_3 at temperatures below 500 C; at higher temperatures oxidation is rapid.[5] The vapor pressure of the oxides is high enough to enable them to vaporize rapidly in vacuum at 1200 C (i.e., in the normal range of heater operating temperatures), causing contamination of the heater coating and other effects which are considered in later sections.

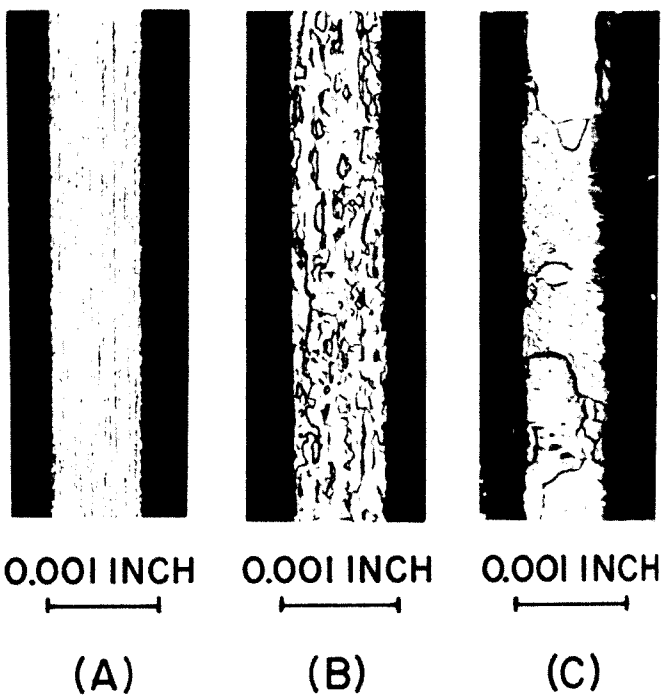


Figure 10. Photomicrographs of Tungsten Wire: (A) as received in fibrous (high strength) state; (B) after flashing to 54 per cent of fusion current, showing onset of recrystallization; (C) after flashing to 63.5 per cent

Dowmo as a Heater Material

The main disadvantage of Dowmo as a heater material is rapid "heater growth" on intermittent operation due to the low hot strength and rapid recrystallization. Heater growth phenomena have been described in detail by J. P. Stelmak,[6] A. G. F. Dingwall,[7] and W. A. Hassett.[8] Growth is most apparent with coiled heaters. If there is sufficient space for a coiled heater to extend, it may "stretch" to twice its original length; tube failure results when the heater touches other parts or comes in contact with and cracks the glass envelope. Such extreme elongation cannot occur with folded heaters, although measurable changes in dimensions do result in some instances. Here, also, folded Dowmo heaters are more subject than tungsten to the "twisted heater"[6] growth effects as discussed in a later section.

The reduced reliability of tubes with Dowmo heaters

associated with their low hot strength and low recrystallization temperature (200 to 300 K lower than tungsten) has resulted in the gradual elimination of Dowmo as a heater material. If, however, the wire is reasonably heavy and the design temperature is low, the net failure rate is usually sufficiently low to permit Dowmo heaters to be used in commercial tubes. The main advantages of Dowmo are:

1. Noise and hum levels caused by electron emission are lower for Dowmo than for tungsten wire. This effect has been conclusively demonstrated both for heater-to-cathode-leakage hum, and for heater-to-grid-emission effects in series-string types. In some critical situations requiring very low noise levels, regular Dowmo heaters with bare apices can be used, while tungsten heaters would require cataphoretically coated apices and insulation coating on all bare-wire areas above the weld to obtain equivalent freedom from noise.
2. Smaller apex chip size on the larger wire diameters due to the greater ductility of Dowmo.
3. Less "spring back" of unannealed heaters of larger wire diameters which makes insertion into the cathode easier.
4. Greater ease of coil winding.
5. Greater ease of welding.

Insulation Coating Materials for Heaters

The high temperatures at which heaters operate place severe requirements on any heater coating material. No organic materials can withstand these high temperatures; among inorganic materials, only certain simple oxides possess the necessary combination of chemical inertness and electrical properties at elevated temperatures.

Now that highly purified materials are available in commercial quantities, it is generally accepted that the best and most stable insulation coatings are obtained by use of the purest possible materials. Twenty years ago, however, it was found advantageous to add small amounts of barium oxide[9] or beryllium oxide[10] or both to the heater coatings to reduce leakage; today, this practice is no longer followed in production. At the earlier period, the aluminum oxide used for heater coatings contained well above one per cent impurities. At that time, there was substantial evidence that additives were frequently effective in reducing heater-cathode leakage. More recent tests[11] with purer materials indicate that if there are any benefits to be gained by adding barium oxide or beryllium oxide, they are, on the average, small. On the other hand, tests show that substantial quantities of certain oxides can be added to the present type of heater coatings without causing a drastic increase in leakage. It is, therefore, conceivable that minor decreases in leakage levels might be obtained through physical effects such as: (1) recrystallization of the insulation grains to give larger particle sizes with reduced areas of contact with the cathode; (2) changes in total emissivity coefficient producing more efficient heat transfer to the cathode and, thus, a lower heater temperature; (3) a change in the porosity of the insulation coating from the present value of 50 per cent of theoretical density; and (4) reduced cracking of the heater coating.

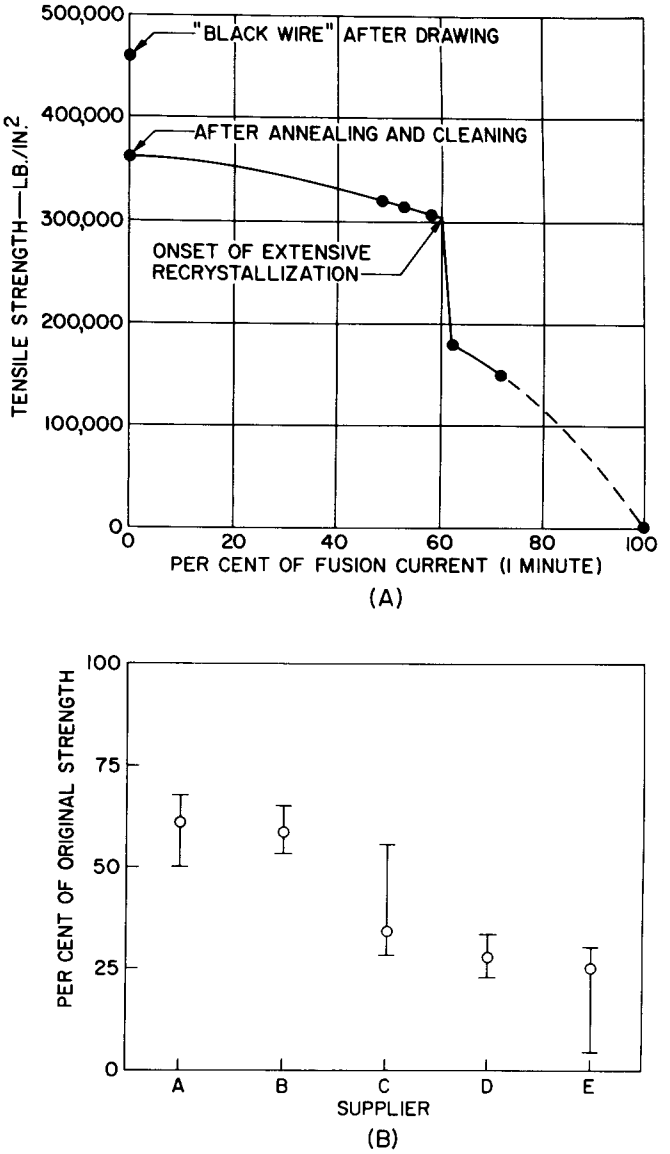


Figure 11. Effect of Temperature on Strength of Tungsten Wire: (A) typical effects of high-temperature on tensile strength of tungsten heater wire; (B) variability of tungsten wire strength after flashing to 62 per cent of fusion current for one minute

Table IV summarizes the properties of the metal oxides that have a melting point in excess of 1800 C, and which might, therefore, be possible heater-coating materials. Examination of Table IV shows that only Al_2O_3 and BeO offer much promise as a basis for heater insulation coatings.[12] High MgO ceramics were used at one time, but at high temperatures on tungsten these decomposed to a sufficient extent to give metallic magnesium deposits on other tube parts. Beryllium oxide has been tried in the past as a heater material with satisfactory results, but it is no longer used mainly because of its toxicity and high cost. Furthermore, a

high melting point and a low bulk density cause technical difficulties in the preparation of dense, sintered insulating coatings of 100 per cent beryllium oxide.

Table IV

Physical Properties of Oxides with Melting Points Greater than 1800 C

Name	Formula	Melting Point Deg C	Properties at Operating Temperatures of Heaters
Aluminum Oxide	Al_2O_3	2045	Stable, excellent electrical insulator
Beryllium Oxide	BeO	2530	Table, excellent electrical insulator, toxic
Calcium Oxide	CaO	2580	Hydrates, emits electrons, poor electrical insulator
Cerium Oxide	CeO_2	2600	Poor electrical insulator
Chromium Oxide	Cr_2O_3	1900	Reduction-sensitive, fair electrical insulator
Hafnium Oxide	HfO_2	2810	Poor electrical insulator
Lanthanium Oxide	La_2O_3	2315	Reacts with H_2O and CO_2, poor electrical insulator
Magnesium Oxide	MgO	2800	Fairly stable, but decomposes at high temperatures on hot tungsten
Nickel Oxide	NiO	1990	Poor electrical insulator
Strontium Oxide	SrO	2430	Hydrates, emits electrons, poor electrical insulator
Thorium Oxide	ThO_2	3050	Poor electrical insulator, reacts on hot tungsten to become emitter
Titanium Oxide	Ti_2O_3	2130	Reduction-sensitive, poor electrical insulator
Vanadium Oxide	V_2O_5	1967	Water-soluble, reduction-sensitive, poor electrical insulator
Yttrium Oxide	Y_2O_3	2410	Poor electrical insulator
Zinc Oxide	ZnO	1800+	Sublimes at 1800 C, poor electrical insulator
Zirconium Oxide	ZrO_2	2715	Reacts on hot tungsten, poor electrical insulator

Aluminum oxide has been used exclusively for many years as the insulating material for heaters. Recent trends have been mainly in the direction of materials of higher purity to give lower heater-cathode leakage.

The Bayer process is generally the starting point of aluminum-oxide manufacture. In this process, the raw material, bauxite, a naturally occurring $Al_2O_3 . 2H_2O$, which contains as much as 25 per cent of Fe_2O_3, SiO_2 and other impurities, is ground and digested in a sodium hydroxide solution at elevated temperatures and under pressure. Insoluble sludges containing the bulk of the Fe_2O_3, SiO_2, and TiO_2 are filtered away from the sodium aluminate solution which is formed, and from which purified aluminum hydrate is subsequently precipitated. The hydrate is then calcined to convert it to the high-temperature form of alumina, α-Al_2O_3, and may, in addition, be acid-washed to remove much of the residual alkalies and iron. Other purification stages in the process are not uncommon.

While the calcined aluminas are satisfactory for heater insulation materials, fused aluminas (e.g., Alundum, a trade name) are now the standard materials used for heater insulations in the electronics industry. The fused aluminas are obtained by fusing the calcined material in an electric arc furnace. Although the fused materials are chemically identical to the calcined aluminas, fusion vaporizes away many contaminants, eliminates firing shrinkage, and produces denser particles of a more desirable angular shape. After fusion, the alumina is ground in an iron or alumina ball mill. The alumina is then separated into particle size ranges, and acid-washed to remove iron contaminants introduced in the milling operation. A typical chemical analysis of the fused material currently used in production is shown in Table V.

Table V

Typical Chemical Analysis of Fused Aluminas Used in Heater Coatings

Material	Percentage Present
Al_2O_3	99.80
Fe_2O_3	0.05
SiO_2	0.03
Na_2O	0.07
Carbon	0.005
MgO	trace
TiO_2	trace
ZrO_2	trace
CaO	trace
$SO_4^=$	trace

Fused aluminas are used in two particle-size grades. The larger-particle-size "500" grade has an average particle size of approximately 19 microns (0.74 mils) with a maximum particle size of approximately 50 microns (2.0 mils). Particles of this size would pass through a standard 500-mesh screen. The finer "900" material has an average particle size of 7.5 microns (0.30 mils) with a maximum particle size of 30 microns (1.2 mils). The larger particles are used almost exclusively when the insulation is applied by drag coating because they permit heavy coatings to be built up quickly. The finer particles are used in spray and cataphoretic coatings because they minimize any settling-out from the suspensions.

APPLICATION OF HEATER COATINGS

Heater coatings are applied by drag coatings, spraying, or cataphoresis. For each type of process, specially tailored suspensions of aluminum oxide are required.

A simple drag-coating process is used in the manufacture of folded heaters currently found in most receiving tubes. The suspension for this process consists basically of 500-mesh fused aluminum oxide in a water-and-methanol solution of aluminum nitrate. In the drag process, a continuous tungsten wire is "dragged" over rollers in contact with the agitated suspension in order to apply a thin layer of coating to the wire. As the wire leaves the rollers, it enters an oven kept at tempera-

tures between 500 and 900 C to dry and bake the coating. The speed of the machine and the viscosity of the suspension are adjusted so that, after approximately one dozen passes, the coating is built up to the proper diameter. The coated wire is then passed briefly through a hydrogen furnace operating at a temperature between 1100 and 1300 C to reduce any oxide produced on the wire during the low-temperature baking. The temperatures used in the drag-coating process are not high enough to sinter the aluminum oxide grains to any appreciable extent; the tough bonding of the finished coating is a result of the aluminum nitrate in the suspensions which decomposes in a complex manner during the baking schedules to an aluminum hydrate and finally to aluminum oxide. The toughness of the applied coatings, therefore, depends to considerable extent on the baking schedules employed.

Spraying techniques are used to apply the insulation coating to most double-helical heaters. The finer grade, 900-mesh, of aluminum oxide, is necessary to obtain the desired coating texture and adherence, and to retard the settling of the aluminum oxide from the suspensions. The spraying suspensions contain a mixture of carefully selected organic solvents to aid in dispersion of the aluminum oxide and to provide the correct vapor-pressure balance for spraying deposits of the correct "wetness," and to give the proper rate of drying. Small percentages of nitrocellulose are also added to give strength to the sprayed coatings so that they can be handled before the firing operation. The coatings are alternately sprayed and dried on a rotary spraying machine until deposits of the correct thickness are built up. After spraying, the heaters are fired in wet hydrogen at 1700 C for a few minutes to sinter the aluminum oxide grains.

Cataphoresis (a form of electrophoresis) is presently used as the means of applying aluminum oxide insulation to single-helical heaters and to the bare apices of folded heaters for premium-type tubes, as well as in the manufacture of spacer coils. By use of appropriate additions of certain soluble inorganic salts (e.g., dehydrated mixtures of aluminum and magnesium nitrates) to aluminum oxide suspensions, a surface charge is produced on the individual aluminum oxide particles. In cataphoresis, the process presently used, the formulations are such that aluminum oxide particles are given a positive charge and, hence, they will be deposited on a heater wire if it is kept at a negative potential in the suspension. Negatively charged particles can, however, be produced by changing the composition of the suspension, but such suspensions have been less efficient. At present, both water and alcohol, or mixtures of organic solvents, are used as the liquid part of the suspensions. The most efficient formulations, which are quite complex, have been developed experimentally.

Very uniform insulation coatings of moderate thickness can be obtained through cataphoresis; difficulties are, however, experienced in producing heavier insulation thicknesses. After heaters or coils are cataphoretically coated with aluminum oxide, they are fired in wet hydrogen at 1700 C for a few minutes to sinter the aluminum oxide.

HEATER BEHAVIOR ON LIFE

Heaters are normally tested under both continuous and intermittent operating conditions as part of routine tube-testing procedures. Intermittent tests for equivalent "lighted" time are more severe than continuous tests because alternate lighting and cooling of the heater mechanically stresses it twice during every cycle. The basic types of failure which can occur are: (1) heater-cathode shorts; (2) shorts between different parts of the heater; (3) "open" heaters (i.e., heaters with broken wires, or open welds); (4) excessive heater-cathode leakage; (5) changing heater current; and (6) heater growth failures.

The reliability of a heater depends on several factors including: (1) type of heater, i.e., coiled or folded; (2) wire diameter and, hence, mechanical strength of the heater wire; (3) heater temperature; (4) polarity and magnitude of any voltage applied between heater and cathode; and (5) the thickness of insulation. Conservatively designed heaters are capable of reliable operation for thousands of hours. Difficulties are generally attributable to well-defined causes which can be corrected.

Failure-Rate Patterns and Heater Temperature

While the polarity and magnitude of the heater-cathode voltage determine to a large extent the possible sources of heater difficulties during life, heater temperature largely determines how soon failure may occur. The easiest manner of assessing the effect of heater temperature on life is by altering the heater voltage during the test [see Eq. (1b)]. According to kinetic theory, heater life might be expected to decrease exponentially with increasing temperature in the manner:

$$L = C \exp - \frac{E}{RT} \qquad (5)$$

where: L = hours to cause a given percentage of failures
T = average heater temperature (deg K) during test
E = activation energy (a constant for a particular lot of heaters, usually about 110 K cal/mol)
R = A universal constant = 1.986
C = a constant depending largely on heater type and testing conditions

A more empirical approach, suitable for rapid calculations of the temperature effect, indicates that the heater failure rate will generally increase approximately as the 12.5th power of the heater voltage. This approximate value has been roughly confirmed in tests completed to date at RCA. According to Eq. (5), a plot of the logarithm of the median life (i.e., the time required for 50 per cent of the heaters in a lot of tubes to fail against the reciprocal of the calculated heater temperature during the test) should result in a linear relationship with a slope corresponding to the activation energy. Fig. 12 illustrates both a verification of Eq. (5) for a series of cycling tests run at different heater

voltages and the approximate validity of the empirical 12.5th-power rule. It may be seen that the median life of a group of heaters decreases rapidly with heater temperature. Consequently, the most generally effective method for increasing heater life is to reduce the operating temperature.

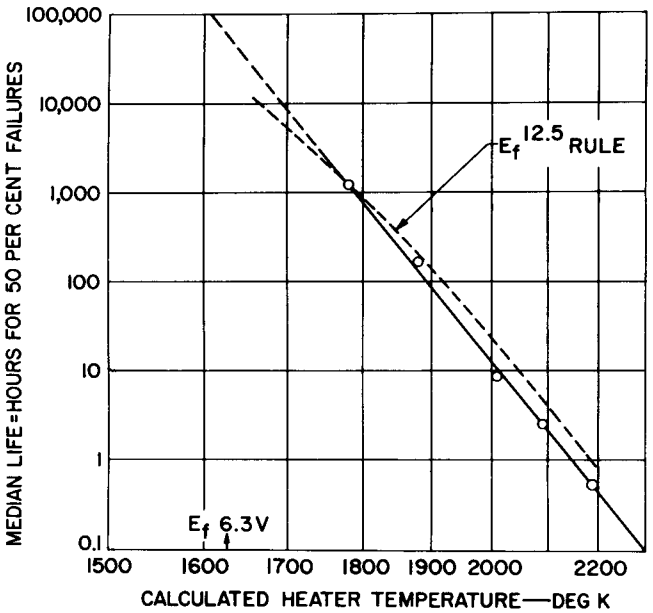


Figure 12. Typical Dependence of Heater Median Life On Heater Temperature (i. e., heater voltage) During a Life Test

The complete pattern of failures for the same lot of tubes considered above is shown in Fig. 13 on graph paper with a cumulative probability of failure scale as the ordinate and the logarithm of the testing time as the abscissa. As can be anticipated from theory, a series of parallel lines provides the best over-all picture of the experimentally observed percentages of tube batches that have failed after any given time of testing. From Fig. 13, it may be concluded that, at least for this particular lot of tubes, the failure distribution was approximately normal with respect to the logarithm of time, and that the failure pattern to be expected at any temperature level could be estimated with some reliability from tests at other temperatures. Similar test results have been obtained on other tube types under substantially different test conditions.

Detailed consideration of such heater-failure patterns indicates that even when failures are all of the same type (as was true in the previous case), there should be an initial period with a high time rate of failure. After this period, the failure rate will decrease. Experimental attempts to improve the heater quality of a lot of tubes by "testing out" the early-hour failures have resulted in some quality improvement among the survivors.

Heater-Cathode Bias

During regular life tests, voltage gradients as high as 20,000 to 40,000 volts per centimeter exist across the heater insulation. At the high temperatures of heater operation, the mobility of many contaminants in aluminum oxide is appreciable and electrolysis through the heater coating may result. With heater-positive biases ($+E_{HK}$), positive-ion contaminants can migrate through the insulation from the heater towards the cathode. For heater-negative biases ($-E_{HK}$), the direction of migration of positive-ion contaminants is from cathode to heater. Because different types of contaminants are introduced into the heater-cathode system with different polarity biases, the possible occurrence of heater-cathode shorts, rising filament current, and heater-cathode leakage on life is strongly sensitive to the heater polarity. Furthermore, preliminary studies indicate that increasing the heater-cathode potential difference increases the rate of the electrolysis effects. If, however, the voltage is raised beyond a critical value (dependent on temperature), instantaneous breakdown of the insulation will occur. Life tests at zero bias or with ac bias have shown features characteristic of both heater-positive and heater-negative tests.

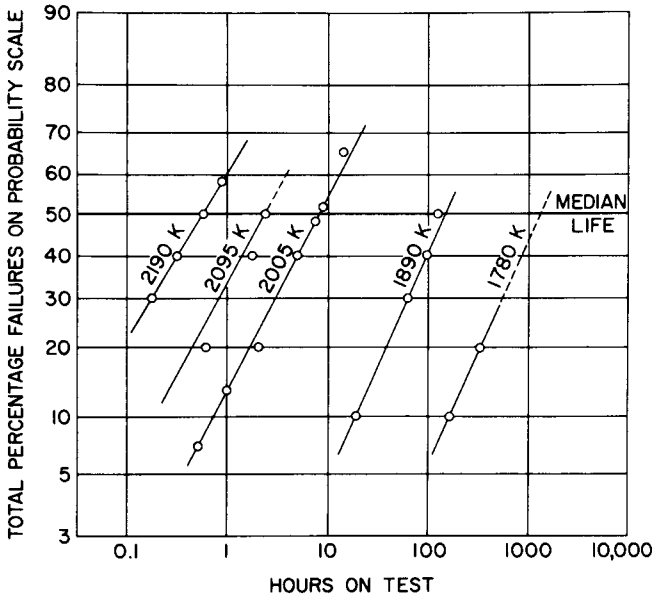


Figure 13. Typical Cumulative Probability Failure Patterns for Heaters Operated at Several Heater Temperatures

Heater-Positive Biases. Heater-positive biases are specified for most life tests because of a reputedly greater severity in terms of the occurrence of heater-cathode shorts — especially under continuous-heater-voltage testing. With this polarity, positive ions will migrate towards the negative cathode. High heater cathode leakage, when caused by positive-ion contaminants in the insulation coating, is usually reduced rapidly because the positive ions are electrolysed to a free surface and vaporized away from the heater-cathode system.

The electrolysis of positive tungsten ions from the heater wire into the heater coating is a potentially undesirable effect of testing with a heater-positive bias.[7,13] The tungsten ions apparently originate from

heater wires which become mildly oxidized in heater manufacture, in tube processing, or during the life test itself. This effect is deduced because tubes showing it can usually be associated with a given production lot of tubes. Once tungsten ions have been electrolysed into the heater coatings, they will either combine chemically with the heater coating or be reduced to metallic tungsten particles, producing not only a visible darkening of the previously white insulation, but also a serious deterioration of the dielectric strength of the insulation. Tube failure may then result through "tungsten crater" heater-cathode shorts or through the effects of rising heater current as the coating is darkened. [7]

The typical tungsten-crater heater-cathode shorts that are associated with heater-positive life have been described in detail by Metson[14] and his co-workers. Their work also suggests that electrolytic decomposition of the aluminum oxide may be an additional factor in the oxidation of the tungsten wire. Three distinct phases were noted during the course of failure: (1) a slight greyish staining of the outer surface of the heater coating at localized areas, (2) an intensification of color at the center of the stains, and finally (3) a crater-type failure evidenced by a small black slagged area on the heater coating. It is characteristic of this type of failure that, in spite of the darkened area, leakage between heater and cathode generally remains low until a few minutes before actual breakdown of the insulation.

The visible darkening of areas of the heater coatings, which is associated with the early stages of tungsten-crater shorts, can, if widespread, cause tube failure through rising heater current. In this case, tube failure may be due to an increase of heater current beyond specifications or, more usually, to the early deterioration of tube characteristics as the cathode becomes overheated because of the additional power supplied by the heater.

In a series of extensive tests to study this problem[7], it was noted that the darkening of the heaters occurs only under heater-positive life test conditions, and proceeds more rapidly at higher temperatures and higher heater-cathode voltages. The increasing heater current is due mainly to the increased thermal emissivity of the visibly darkened coating which improves the efficiency of heat transfer between heater and cathode. In severe cases, the blackening of the heater coating may be sufficient to cause the heater current to rise by 10 to 20 per cent in the course of 1000 hours under normal testing conditions. Statistically designed experiments have shown that this problem can be largely overcome by use of suitable aging schedules. An alternate possibility, of course, would be to reduce the operating temperature of the heater to slow down the rate at which these effects occur.

Heater-Negative Biases. With heater-negative biases, the previously considered darkening of the insulation coating does not occur, and the heaters retain their original white appearance throughout life. On the other hand, electrolysis of metallic ions into the heater coating from the cathode may result, causing increasing heater-cathode leakage and embrittlement of the heater wire or both. A third effect, which occurs on cycling tests, is an increased rate of heater growth.

Increasing heater-cathode leakage on life, especially $+I_{HK}$ (leakage measured with the heater positive) is a typical symptom of heater-negative or "zero" bias conditions. Gas bombardment of the cathode with sufficient intensity to splatter cathode material on the heater can, however, produce similar symptoms, but this occurrence may be readily detected by examination of the cathode. Life tests at RCA and elsewhere have shown that the possible development of $+I_{HK}$ leakage occurs most rapidly when the heater is operated under a high negative bias voltage. This high negative bias causes an increased rate of contamination of the heater coating by electrolysis of the more active additives of the cathode base metal into the insulation. The $+I_{HK}$ leakage which develops under zero-bias life conditions will increase less rapidly than under negative-bias conditions because electrolysis does not occur. Finally, under heater positive tests, the electrolysis is in the reverse direction and leakage is usually maintained at low levels. Fortunately, since the highest heater-cathode leakage tends to occur with a "direction" opposite to that of the test, the polarity of the heater and cathode need to be interchanged before it can be detected. This unidirectional leakage effect is illustrated in Table VI. The data indicate that leakage occurred because of the accumulation of contaminants in the insulation. The contaminants produce leakage unless actively electrolysed away from the cathode. Considerable success in reducing the development of such leakage on heater-negative or zero-bias life tests has been obtained by use of cathodes with low levels of magnesium.

A similar effect, which has been reported to occur under heater-negative conditions, is the electrolysis of nickel from the cathode onto the heater wire. [15] Nickel, when present in even minute amounts, will severely embrittle tungsten. As a result, the ability of a heater wire to withstand intermittent operation without breaking could be substantially reduced.

The greatest severity of heater-negative biases, however, usually occurs on intermittent tests due to an increased rate of heater growth. While heater growth will also occur with zero or heater-positive biases, it is apparently accelerated by a factor of two or three under the heater-negative conditions. [6, 7] Such heater growth occurs when a heater is alternately heated and cooled, and is unable to expand and contract freely within the cathode. As a result, the heater becomes progressively distorted. In coiled heaters, the coil can stretch to a clearly visible extent. In the most severe cases, where Dowmo wire has been used, heater coils have stretched to nearly twice their original length and have caused tube failure after touching the glass bulb or other components. The corresponding effect in folded heaters, is that the once straight heater legs assume a curly or twisted shape. Failures, in this case, may eventually occur when bare heater apices come in contact with one another or with the cathode. The thermal stressing of the intertwined folded heater legs, which cannot expand and contract freely as the heater is cycled, may also produce substantial numbers of broken heater wire failures, especially after the wire has become embrittled through recrystallization or

nickel contamination. Reduction of the heater design temperature to below 1550 K has generally eliminated such complaints.

Table VI
A comparison* of Heater-Cathode Leakage Levels Resulting from Heater-Positive, Zero-Bias, and Heater Negative Operation to Illustrate Electrolysis Effects.†

Hours on Regular Life Test	Heater at +50 volts throughout Life		Heater at zero volts throughout Life		Heater at -50 volts throughout Life	
	$+I_{HK}$	$-I_{HK}$	$+I_{HK}$	$-I_{HK}$	$+I_{HK}$	$-I_{HK}$
0	<0.1 μa	<0.1 μa	<0.1 μa	<0.1 μa	<0.1 μa	<0.1 μa
1000	110	1.7	97	2.2	8.2	2.0
2500	170	1.8	150	2.7	11.5	4.0

*The rate of leakage development illustrated here is unusually high

†The leakage measurements, $+I_{HK}$ and $-I_{HK}$, were made at various down periods during life with the heater 100 volts positive and 100 volts negative respectively. The tube tested was type 5751

HEATER-CATHODE LEAKAGE

Definition and Characteristics

Heater-cathode leakage is an undesired flow of current between heater and cathode in spite of the insulation coating on the heater wire. Although the currents involved are relatively small (usually of the range 0.01 to 100 microamperes) at normal operating temperatures, they can produce undesirable effects on circuit performance[16] so that measurement of this leakage current is standard for rating almost all tube types.

The size of the leakage current depends on the magnitude and polarity of the voltage between heater and cathode. It is customary to report the leakage values measured when the heater is 100 volts positive with respect to the cathode ($+I_{HK}$) or 100 volts negative with respect to the cathode ($-I_{HK}$). Fig. 14 shows two typical heater-cathode leakage characteristics. In one case, the heater-positive ($+I_{HK}$) leakage is the higher, and in the other case the heater-negative ($-I_{HK}$) leakage is the higher. It is quite typical, however, that in neither case are the "positive" and "negative" leakage currents equal nor does the leakage current increase linearly with increasing voltage as would be expected from an impedance obeying Ohm's law. Although the values of $+I_{HK}$ and $-I_{HK}$ tend to average out to about the same levels on tubes with commercially-low leakage levels, it is usually characteristic that in "high-leakage" lots of tubes the $+I_{HK}$ values predominate while in "very low-leakage" lots, the $-I_{HK}$ values more often predominate. A third way in which the heater-to-cathode-leakage characteristic differs from that of an "Ohm's law resistor" is that the leakage current will not generally fall to zero when zero bias is applied between heater and cathode and, as a result, the heater cathode impedance becomes small at very low biases. Such effects are important in heater hum; in critical low-level circuits, the heater is frequently operated at a level 20 to 50 volts above cathode potential to avoid the zero-bias region. Studies of this effect by pulse measurements on unipotential heaters indicate that it is only partially due to the applied heater voltage which supplies thermal energy to the heater wire. It may be noted that if electron emission contributed to leakage, the current would not be expected to become zero at zero applied bias. This observation suggests that electron emission is a factor in heater-cathode leakage.

Decay and Polarization

When a voltage is applied between heater and cathode, initial "surges" of leakage current are observed which invariably decay to lower levels with time. After the voltages are removed, the heater cathode system recovers to its previous state, and the decay and recovery cycle may be repeated. Such temporary leakage decay is called polarization.

Polarization is a typical symptom of systems in which conduction of current is associated with the presence of charged carriers having appreciable mobility. The effect of an applied voltage is to cause a drift of the charged carriers. Consequently, complex voltage gradients, which are dependent upon the distribution and number of the carriers at any time are developed. At equilibrium, most of the voltage drop will normally occur in the vicinity of an electrode and, thus, only the remaining fraction of the originally applied voltage is effective in promoting current flow through the rest of the system. Consequently, the equilibrium current level can be substantially below initial levels.

Current decay and complex voltage gradients throughout the Al_2O_3 insulation coating are observed in heater-cathode leakage studies[17] indicating that ionic contaminants play a major role in leakage. One noteworthy consequence of the effects of polarization is that most of the static heater-cathode impedance is developed at the boundaries of the insulation coating and, therefore, increased insulation thicknesses would not be expected to produce proportionately increased resistance. On the contrary, the most noticeable effect of an increased amount of insulation can be that it introduces a greater total amount of leakage-producing contaminants into the heater-cathode system and, therefore, may require substantially longer processing to achieve equivalently low leakage levels.

Measurements of heater-cathode leakage at various times show both rapid-decay time constants (milliseconds) and slow-decay time constants (minutes). At normal heater temperatures, much of the loss is recoverable although the over-all leakage level may be continually reduced during prolonged measurements since contaminants are vaporized out of the system. The extent of the decay period and the amount of initial leakage current will depend markedly on the state of the heater-cathode system before measurement. For example, if the heater-cathode bias were suddenly changed from a high negative value to a high positive value, an unusually large initial leakage current surge would be noted. Due to such effects, it is standard practice in testing tubes to preheat tubes for several minutes with the same heater-cathode bias that is to be used during measurement of the leakage current. In

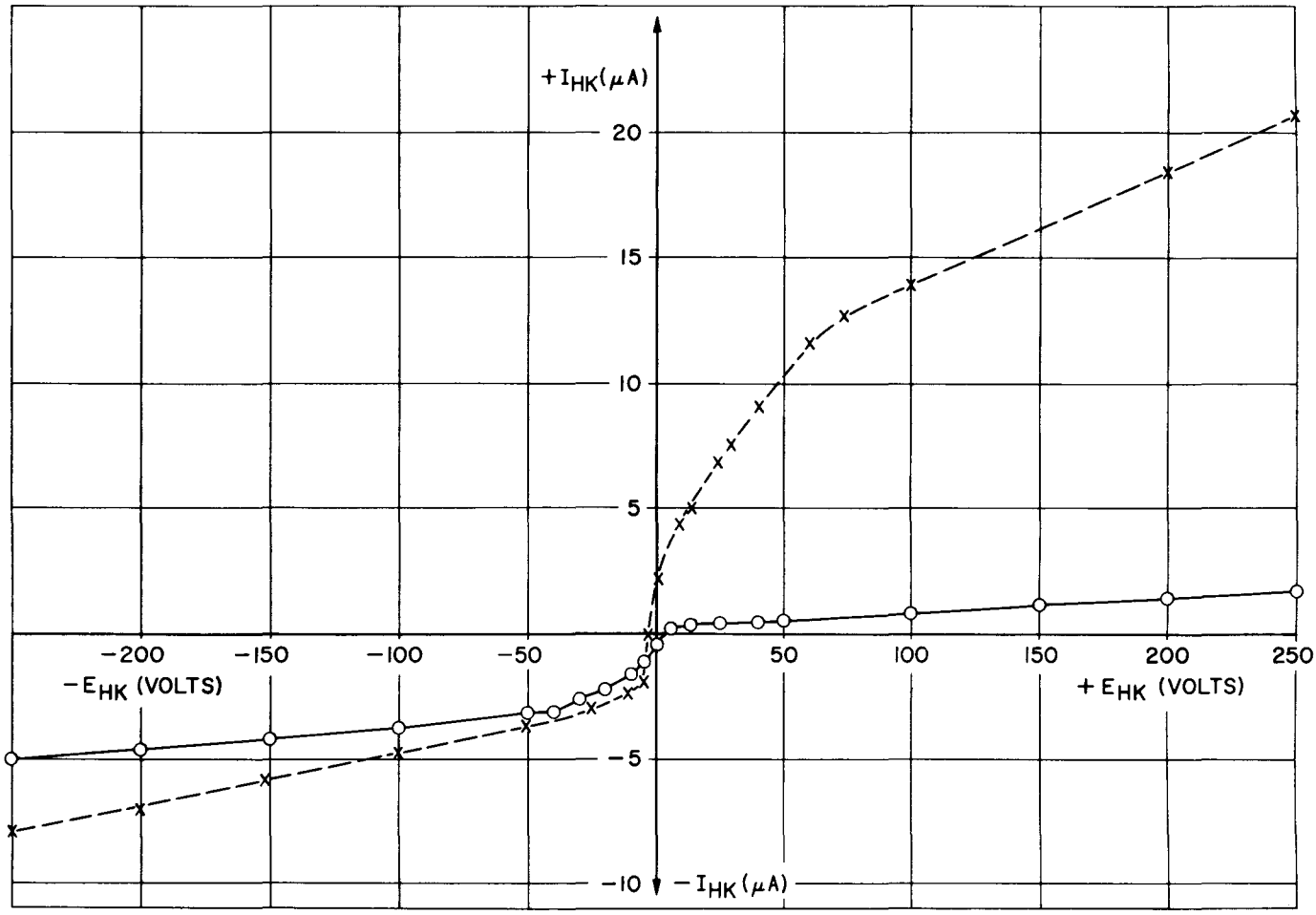


Figure 14. Typical Heater-Cathode Leakage Characteristics on Two Different Tubes by Conventional DC Methods of Measurement

this way, much of the transient leakage current can be eliminated and, consequently, reasonably stable meter readings can be obtained.

The decay effects associated with heater-cathode leakage can have an important influence on leakage measured under different conditions. Fig. 15 shows the heater-cathode impedance of a tube measured at different frequencies. The impedance is substantially lower at the higher frequencies since the rapidly reversing polarities accentuate polarization effects. Fig. 16 shows the leakage measured on a tube by two techniques. The upper dashed-line curve gives the heater-to-cathode-leakage pattern when a 60-cycle sweep with increasing voltage is applied between heater and cathode. The lower dashed-line curve is for decreasing sweep voltage; here, distinct evidences of hysteresis caused by leakage decay are observable. The upper, solid-line, curve shows the peak leakage measured by pulse techniques (1 millisecond, 60 pulses per second).

Distribution of Heater-Cathode Leakage Values

Heater-cathode leakage values invariably occur in highly skewed statistical distribution patterns. It is characteristic that extreme variations are found in "identically" processed tubes; leakage values varying by a ratio of 100 to 1 are not uncommon in measurements on a large lot of tubes.

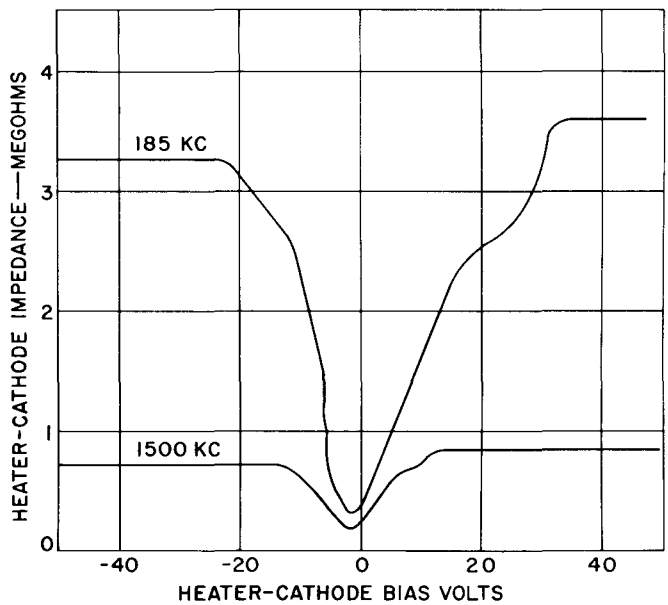


Figure 15. Typical Frequency Effects on Heater-Cathode Impedance

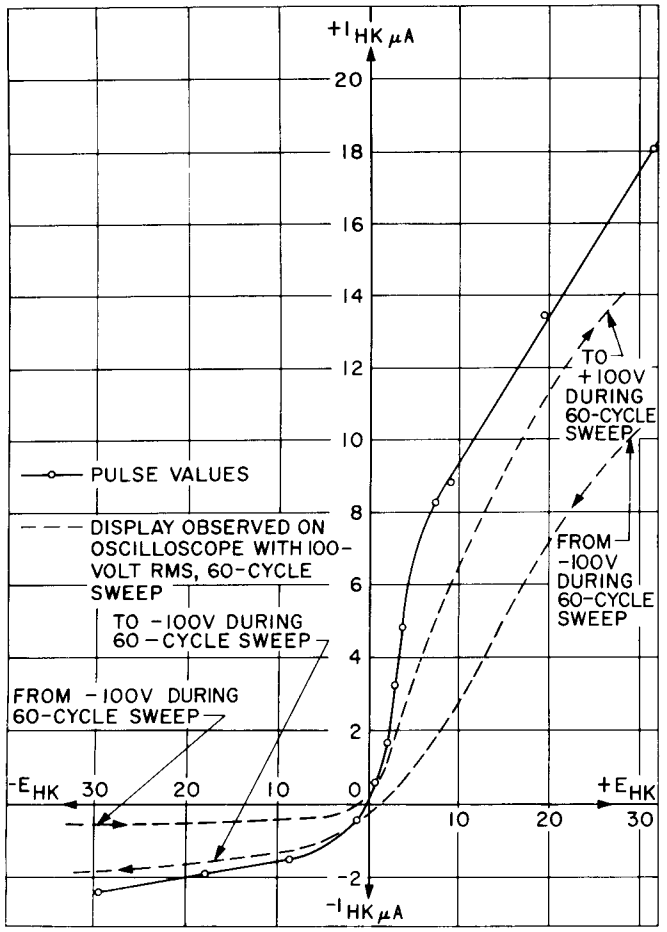


Figure 16. Heater-Cathode Leakage Characteristics Measured by Different Methods on the Same Tube to Illustrate the Effects of Polarization

Because of the highly skewed distribution, a conventional statistical treatment of leakage data, which presupposes a normal, Gaussian distribution, may not be satisfactory. For example, one "high" leakage value might outweigh twenty low leakage values, if straight arithmetic averages were used. From a purely statistical point of view, a more satisfactory, simple method of handling such highly skewed distributions would be to take the logarithms of all leakage data and then to analyse these logarithms in the conventional manner (i.e., to assume a logarithmo-normal distribution). In this way, any single "wild" measurement will not receive excessive emphasis. In this type of statistics, the important parameter is not the average, but the median (the leakage level above and below which, ideally, 50 per cent of the measurements fall). The need for adopting such a logarithmic distribution is easily justified theoretically if it is assumed that the assignable causes of variation are in some manner connected with temperature variability. Since chemical reaction rates depend logarithmically upon temperature, it would follow that a logarithmic distribution of leakages could then result from slight temperature variations.

In Fig. 17, the applicability of a logarithmic distribution of leakage values to experimental data is indicated. This plot is made on cumulative probability graph paper and shows an actual distribution of heater-to-cathode-leakage measurements on both linear and logarithmic scales. The straight-line relationship for data plotted on the logarithmic scale indicates that a logarithmo-normal distribution provides a reasonable fit to the leakage data.

Changes of Leakage after Tapping of Tube

Changes in heater-cathode resistance on a small percentage of a lot of tubes after drop-testing, vibration, tapping, cycling the heater voltage on and off, or even after normal handling and storage is one of the major heater-cathode-leakage problems.[11] Tests indicate that the leakage in any tube can be altered after sufficiently severe tapping; it is usually possible to increase and decrease the leakage alternately by locating appropriate points of impact. Fortunately, leakage produced in such a manner will normally disappear after a few hours of operation at normal filament voltages or after a shorter period at increased voltages.

Heater movement within the cathode is without doubt the cause of such variable leakage values. Factory experience indicates that such effects are greatest with tube constructions having short or flat cathodes, loosely fitting heaters, or whenever opportunities for heater movement exist. Conversely, these effects are minimized, for example, by the use of rigid mounts and inverted pinched cathodes which restrict movement.

In tube types where $+I_{HK}$ leakage has sporadically appeared after normal storage, some success in maintaining uniformly low-leakage levels has been achieved by use of a severe hotshot stage near the end of the aging schedule. Such a high-temperature treatment will tend to vaporize contaminants from all parts of the heater insulation, as contrasted with normal aging schedules where contaminants may be thoroughly electrolysed away only from those areas which have been in contact with the cathode. A possible confirmation of the latter effect may be inferred from tests in which it was found that cycling heaters on and off during aging to cause artificial heater movement could considerably reduce the percentage of tubes which subsequently would develop leakage. In addition, such high-temperature treatment results in the vaporization of a visible layer of insulation onto the inner surface of the cathode. Although the importance of such thin insulating films on the metallic surfaces of the heater-cathode system is difficult to evaluate accurately, it is possible that they are the most important factor in achieving high heater-cathode resistances. In this sense, it may be noted that although hydrogen-fired heaters or heaters removed from low-leakage tubes usually have lower leakage values immediately after sealexing than unprocessed control heaters, the leakage values are rarely low enough to be commercially acceptable without additional aging of the tubes. The missing factor could well be that such thin insulating layers are not formed.

Effect of Heater Temperature on Leakage

A temporary increase (or decrease) in heater temperature caused by a change in heater power will al-

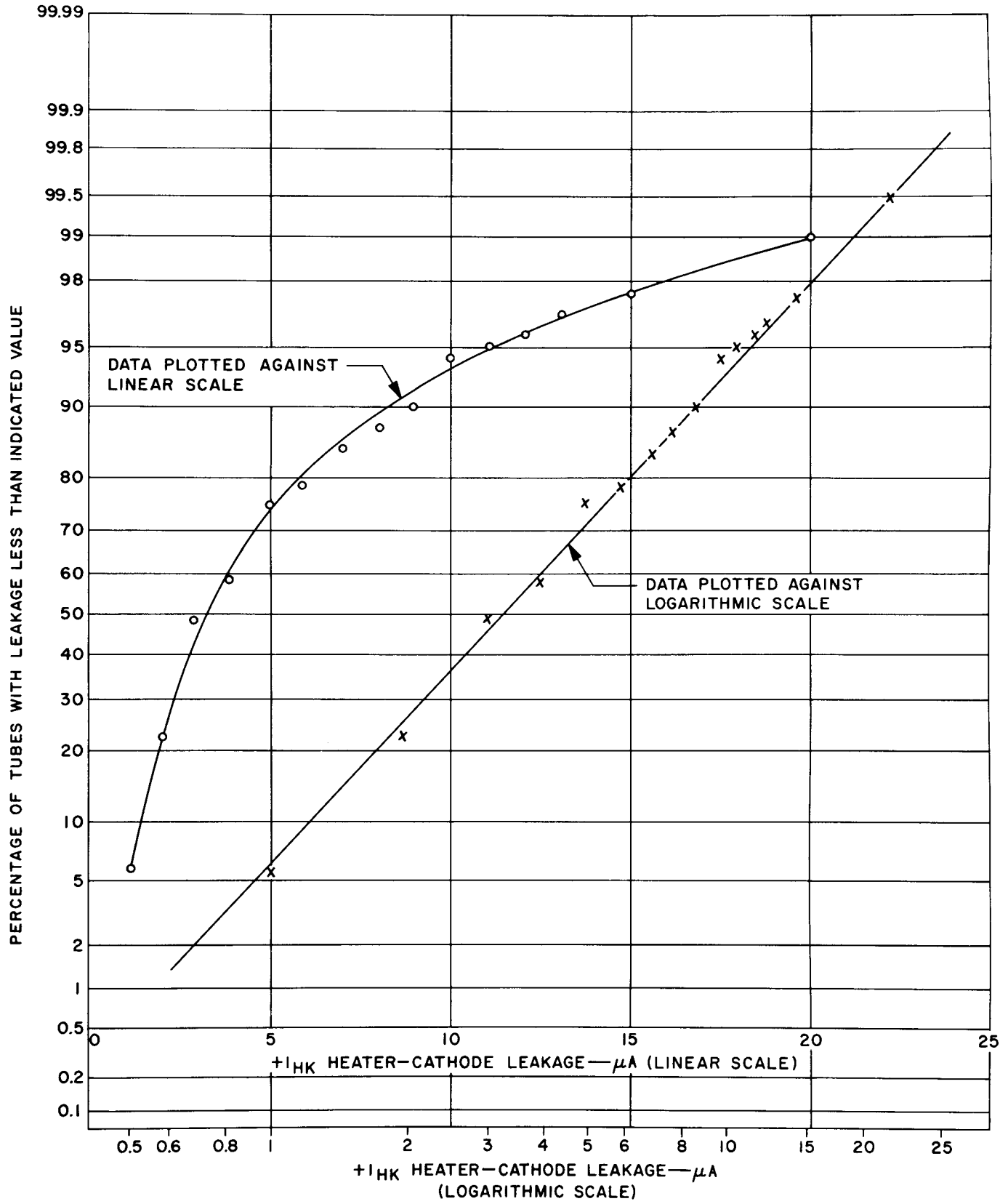


Figure 17. Distribution of Heater-Cathode Leakage Values ($+I_{HK}$) on Cumulative Probability Graph to Show That the Logarithms of Leakage Have an Approximately Normal Distribution (120 tubes)

ways result in a temporary increase (or decrease) in heater-cathode leakage. As shown in Fig. 18, such leakage changes exponentially with the reciprocal of absolute temperature in a manner characteristic both of conduction and emission phenomena.

mensions but does not change the total rated power dissipation of the heater.

The data in Fig. 18 show that the temperature dependence of the $-I_{HK}$ leakage (measured with the heater

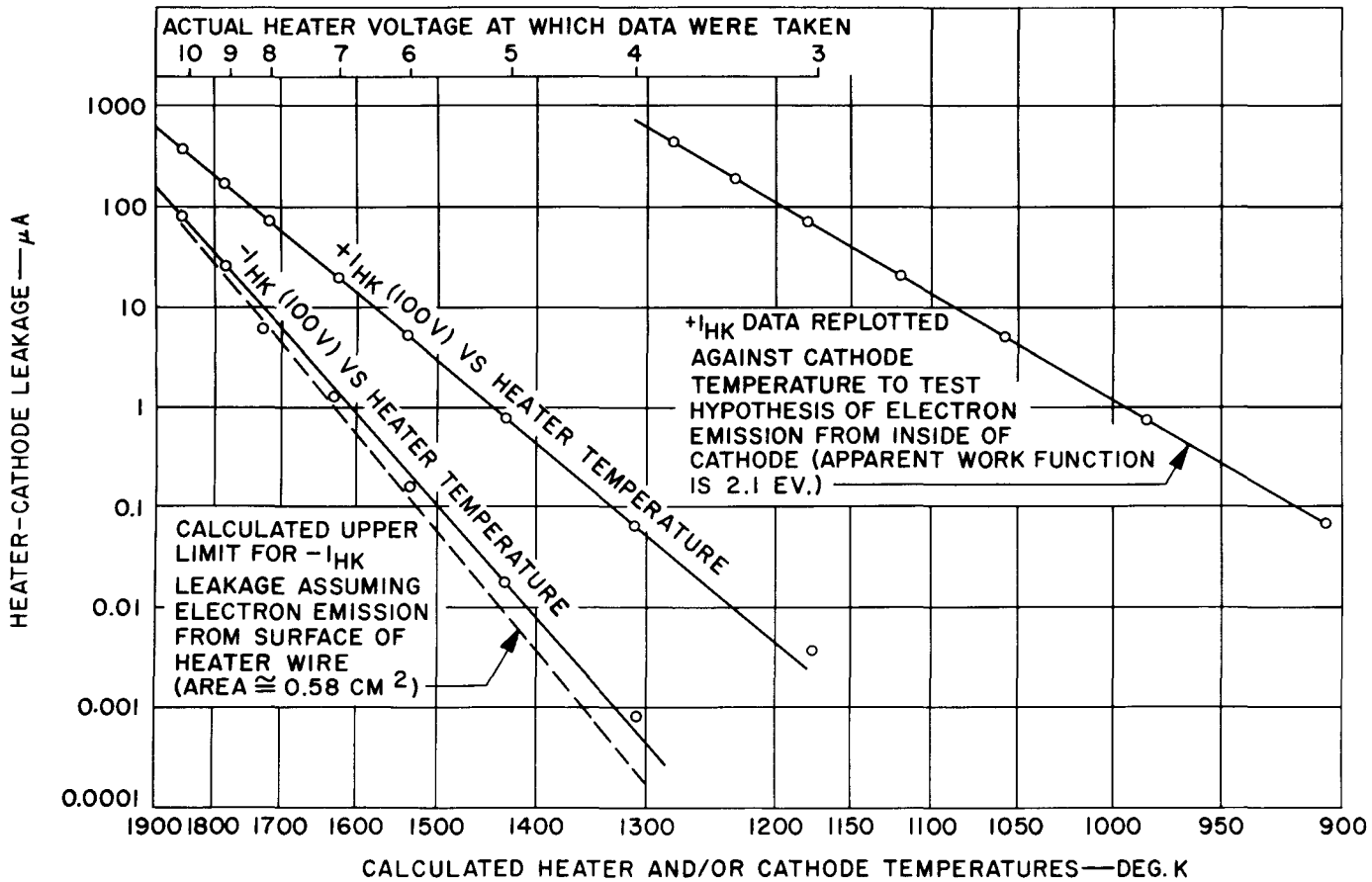

ACTUAL HEATER VOLTAGE AT WHICH DATA WERE TAKEN
10 9 8 7 6 5 4 3
HEATER-CATHODE LEAKAGE — μA
1000
100
10
1
0.1
0.01
0.001
0.0001
+IHK (100V) VS HEATER TEMPERATURE
-IHK (100V) VS HEATER TEMPERATURE
+IHK DATA REPLOTTED AGAINST CATHODE TEMPERATURE TO TEST HYPOTHESIS OF ELECTRON EMISSION FROM INSIDE OF CATHODE (APPARENT WORK FUNCTION IS 2.1 EV.)
CALCULATED UPPER LIMIT FOR -IHK LEAKAGE ASSUMING ELECTRON EMISSION FROM SURFACE OF HEATER WIRE (AREA ≅ 0.58 CM 2)
1900 1800 1700 1600 1500 1400 1300 1200 1150 1100 1050 1000 950 900
CALCULATED HEATER AND/OR CATHODE TEMPERATURES—DEG. K

Figure 18. Typical Dependence of Heater-Cathode Leakage on Heater Temperature Resulting from Changes in Heater Voltage

It cannot generally be inferred from such a temperature dependence, however, that a normally "hot" heater will always have greater heater-cathode leakage than a corresponding normally "cool" heater. This could only be concluded if commercial leakage levels were so low that the ultimate capability of the heater-cathode system were being approached. In the more usual case, commercial leakage levels are well above the ultimate levels because of contaminants in the system. Over the limited temperature range normally used in heater operation, contamination levels are apparently of greater importance than actual heater temperatures in determining leakage levels. As indicated under the subsequent discussion of heater size, it is easier to age out leakage of the type caused by contamination from a small "hot" heater than from a corresponding colder but, of necessity, larger heater. A further factor to be considered is that, regardless of the heater-wire temperature, so long as the power dissipated by the heater remains fixed, the cathode temperature will not change. Thus, any factor in heater cathode leakage which depends only upon the cathode temperature will not be basically altered by a change in heater design which merely changes the heater di-

negative) and the $+I_{HK}$ leakage (measured with the heater positive) is not the same. This observation provides strong evidence that the two types of leakage do not result from the same effect. The simplest explanation is that in the former case, the leakage was dependent upon electron emission from the tungsten heater wire, while in the latter case the leakage was due to electron emission from the inside surface of the cathode.

The order of magnitude of the $-I_{HK}$ heater cathode leakage, which could be due to electron emission from the tungsten heater wire, is readily obtained from heater dimensions and the known electron emissive properties of tungsten. A comparison of this type is shown in Fig. 18; the correspondence between measured and calculated leakage levels for this set of data over a wide temperature range is striking.

In a similar manner, the order of magnitude of the $+I_{HK}$ leakage may be explained under the assumption that it is produced by electron emission from the inside of the cathode. In this case, however, the $+I_{HK}$ data must be plotted against cathode temperature,

rather than heater temperature. This type of comparison is also made in Fig. 18. From the observed temperature dependence, an estimate of approximately 2.1 ev for the work function of the inside of the cathode would be necessary. Work functions of contaminated nickel in the range of 2.0 to 3.0 ev appear reasonable, and although exhaustive information is not available, it appears that the observed orders of magnitude of $+I_{HK}$ leakage over a wide temperature range can be explained in terms of electron emission from limited areas of the inside of the cathode.

The Effects of Heater Size

Although heater-cathode leakage can be controlled satisfactorily in most instances, certain tube types have chronic difficulties, e.g., hum problems. In some of these cases, difficulty may be due to very exacting leakage current specifications. In many instances, however, the susceptibility of a tube type towards leakage problems is largely a matter of heater dimensions.

Generally, the larger a heater, the greater will be the difficulties experienced in "aging out" leakage. This effect is partially due to the greater surface areas of the large heaters across which conduction and emission can occur. An even more important cause of this effect, as discussed previously, is a larger volume, and, hence, the correspondingly larger total amount of possible contaminants in the heater-cathode system. A third important factor is that it is more readily possible to age a normally "hot" heater at high temperatures and, therefore, to achieve efficient sublimation of contaminants from the insulation without recontamination from an overheated cathode or other overheated tube elements.

Experimentally, it appears that leakage susceptibility increases exponentially with heater size. Heater size will increase with an increase in either the voltage or current rating of the heater. Table VII shows a typical correlation between leakage current and heater current rating (i.e. heater size).

In a similar manner, heater volume and susceptibility toward leakage problems increase with heater voltage rating, as is indicated in Table VIII, for a single tube type family differing only in heater voltage rating.

By using a smaller, "hotter" heater, in a given tube type, it has often proved possible to take advantage of the heater size vs. leakage susceptibility relationship. Although substantial improvements in heater-cathode leakage values can frequently be achieved in this manner, the result is obtained at the expense of heater reliability. Improved aging techniques are, therefore, a generally more desirable means of accomplishing the same result.

The Mechanisms of Heater-Cathode Leakage

Aside from the normally negligible energy transfer associated with capacitance, heater-cathode leakage can only be caused by the transport of charged carriers — either ions or electrons — between heater and cathode under an applied electric field. Such charged carriers may proceed through the insulation by conduction and across the vacuum in a tube by emission. Although the relative values of leakage current to be anticipated at any given voltage between heater and cathode depend on whether emission or conduction phenomena predominate, the direction in which the charged carriers migrate depends only upon "carrier charge" and the polarity of the applied field. The following possibilities for the direction of carrier migration are, therefore, complete:

1. $+I_{HK}$ leakage (heater positive): This can result from the flow of positive ions from heater to cathode, or negative ions and electrons, or both, from cathode to heater.

2. $-I_{HK}$ leakage (heater negative): This can result from the flow of negative ions or electrons, or both, from heater to cathode, or positive ions from cathode to heater.

A knowledge of the means by which current passes

Table VII
The Relation Between Heater Design and "Maximum" Heater-Cathode Leakage Reported in a Group of Tubes
(Source: Advisory Group on Electron Tubes[16])

Tube Type	E_f Volts	I_f Amperes	Max. Heater-Cathode Leakage μa	Best fit from Volume Effect * μa
5899	6.3	0.150	5	9
6100	"	0.150	10	9
5670	"	0.175	10	10
5725	"	0.175	10	10
5750	"	0.300	10	13
5726	"	0.300	10	13
5902	"	0.450	15	18
5727	"	0.450	15	18
6005	"	0.600	50	30
6098	"	1.200	75	75

*Best fit to leakage data assuming leakage increases exponentially with heater volume.

Table VIII

The Relation Between Heater Design and Average Maximum Heater-Cathode Leakage in a 10-Tube Sample

(Source: RCA Rating Laboratory Data)

Tube Type	Rated Heater Volts	Av. Max $+I_{HK}$ μa	Best Fit From Volume Effect μa	Av. Max $-I_{HK}$ μa	Best Fit from Volume Effect μa
6L6G	6.3	4.1	5.0	5.4	5.3
12L6GT	12.6	7.4	5.9	12.1	5.8
25L6G	25	8.0	8.4	5.6	7.0
35L6GT	35	15.5	14.0	11.9	10.5
50L6GT	50	31.2	31.2	19.0	14.0

between heater and cathode would indicate what steps, if any, could be taken to increase heater-cathode resistance. The most direct method of determining what effects control a particular instance of heater-cathode leakage would be a study of the characteristic of the leakage vs. the applied heater-cathode bias. Unfortunately, although conventional, static dc leakage measurements are useful as a measure of a tube's performance in actual operation, they are not very useful in evaluating the various components involved in heater-cathode leakage. Polarization, and the consequent formation of complex voltage gradients across the insulation, make it impossible to specify the effective voltages at the surfaces between which leakage occurs. Furthermore, with static-type measurements, the heater is not unipotential at the time of measurement, since it is heated electrically—a situation that confounds leakage measurements at those small biases which are of importance in establishing the existence of emission effects.

Pulse techniques have been adopted in the RCA laboratory as an improved means of studying leakage in detail, because when only small pulses of current are drawn at infrequent intervals, polarization effects can be practically eliminated. Furthermore, the pulse leakage measurements can be made when the heater is unipotential by supplying it with half-wave rectified power and taking measurements during the off half-cycle. Fig. 19 A shows typical $+I_{HK}$ pulse measurements on a tube that has rather high leakage. The net leakage characteristic at small biases (after "subtracting" the linear contributions readily apparent at the higher voltages) is shown on an expanded scale in Fig. 19 B. An interpretation of the measurements in Fig. 19 B may be made assuming electron emission from the inside of the cathode. At the low current levels involved in heater-cathode leakage (microamperes per square centimeter) and the close spacings in the heater-cathode system (of the order of 0.001 inch), the leakage current, if due to electron emission, would show only retarding-field and saturated characteristics. The theoretical type of voltage-vs-current relationship to be anticipated, assuming electron emission from the inside of the cathode to the heater wire is indicated in Fig. 19 B so that comparison may be made with the measured leakage values. The observed onset of "saturation" at a bias voltage of approximately 2.0 to 2.5 volts is in the range to be anticipated for electron emission to a tungsten collector from a contaminated surface having a total work function of 2.0 to 2.6 ev (the values necessary to explain the temperature dependence of $+I_{HK}$ leakage shown in Fig. 18).

Ion emission from the heater coating to the cathode may in theory also occur and will produce leakage-vs-bias voltage characteristics similar to those shown in Fig. 18. The presence of copious amounts of positive- and negative-ion emission currents has been verified both at RCA and elsewhere[18] with experimental structures in which a previously unprocessed heater was physically isolated from the cathode. The observed ion-emission currents, however, were "aged out" rapidly to levels well below those normally encountered in practical problems. It, therefore, appears that ion emission is not an important factor in "well-aged" tubes.

The approximately linear increase in heater-cathode leakage under pulse measurement conditions shown in Fig. 19 A at pulse voltages higher than ten volts may be interpreted in terms of pure conduction through the insulation coating, or pure emission effects, or a combination of the two. Conduction of current may, of course, occur through the aluminum oxide coating at the points at which the insulation touches the cathode. In this event, the conduction and emission phenomena must be viewed as separate effects which occur simultaneously and essentially independently; both would be enhanced by contaminants in the heater-cathode system. On the other hand, the approximately linear increase in leakage at high bias voltages could be explained by the well-known Schottky effect of field-enhanced emission. In the latter case, it may be necessary to assume that the collector is not the tungsten wire, but the surface of the insulation coating in order to account for the intense voltage gradients necessary to have theory and experiment coincide. In both cases, complex voltage gradients can be developed in the insulation coating as a result of current flow through the insulation producing the symptoms of polarization described previously.

Although the role of electron emission in heater-cathode leakage appears well documented, it does not appear possible at present to eliminate electrical conduction through the insulation coating as a possible ef-

fect of camparable magnitude which occurs simultaneously. The most serious argument against a conduction mechanism is that the temperature dependence of $+I_{HK}$ leakage and $-I_{HK}$ leakage is not normally the same. On the other hand, it appears difficult to explain the observed changes in heater-cathode leakage as the result of heater movement unless a conduction mechanism is assumed. It is doubtful, however, that the ultimate intrinsic conductivity of the insulation coating can be an important factor; otherwise it would not be possible to produce any tubes with the very low leakage levels which it is possible to achieve. On the other hand, if conduction type leakage were attributable to ion currents alone, it should be possible to equate the total observed leakage in coulombs to the known amounts of ionic contaminants in the heater-cathode system. All comparisons of this type show, however, many orders of magnitude more total leakage current than could be accounted for on the basis of electrolysis and known amounts of contaminants. It, therefore, appears necessary to assume that leakage is the result of contaminants which promote both electron emission and extrinsic conductivity caused by introduction of impurity levels in the insulation; leakage can persist until the contaminants are vaporized from the heater-cathode system. Some evidence in support of this hypothesis was gained from tests in which the upper end of the cathode was scaled to reduce loss of contaminants from the heater-cathode system; in this instance, leakage levels were significantly higher after aging than the control cathodes having loosely pinched upper ends.

In summary, studies of heater-cathode leakage suggest that both electron emission and electron conduction contribute to current flow, and that the observed leakage can be considered as the summation of the two effects. Both effects would be expected, on theoretical grounds, to be strongly sensitive to the presence of contaminants in the heater-cathode system. In order to minimize leakage, it would, therefore, be necessary to use materials free of leakage-producing, ionic contaminants, to reduce the area of contact between heater and cathode to a minimum, and possibly, to cover with an insulating film all hot metallic surfaces from which electron emission might result.

Factors Contributing to Heater-Cathode Leakage

Practical studies to determine what factors actually cause leakage tend to confirm the implications of theoretical studies. Statistically analyzed experiments have shown that a large number of factors can be directly associated with heater-cathode leakage and hum. Some of these factors are considerably more important than others, or may have a relatively important effect in one tube type, while in other tube types, the same factors may be relatively insignificant. The difference in importance will depend on leakage levels desired, heater size, and aging schedules.

Two factors appear to be of basic importance in considerations of heater-cathode leakage: (1) the presence of certain contaminants in the heater-cathode system; and (2) the type of contact which the heater makes with the cathode surface. The latter is without doubt one of the critical parameters in heater-cathode insulation. Although the main body of the heater operates at about 400 to 500 K above cathode temperature, it seems probable that the temperature of the insulation at the areas of contact must approach cathode temperature. According to Fig. 18, a temperature difference of 400 K results in a thousandfold change in leakage level. It is, therefore, clear that most of the insulating capability of the coating must be developed at the low-temperature points of contact. It is noteworthy, however, that the intrinsic conductivity of the "pure" insulating material in granular form does not appear to be an essential factor. If it were, it would not be possible to produce even a few tubes with leakage levels below 0.01 microampere. For the same reason, the additions of certain of the more stable oxides to the aluminum oxide insulation or decreased coating thicknesses need not cause deleterious increases in leakage. It is also noteworthy that actual contact of the heater with the cathode is not necessary to cause leakage because emission effects are an important factor.

Laboratory tests to assess the exact effects of contaminants on leakage are difficult to perform because of : (1) the large number of uncontrollable variables arising when heaters are made at different times, (2) difficulties in making reliable analyses of contaminants present to the extent of a few one-hundredths of a per cent by weight in chemically-inert materials, and (3) the very wide range of leakage values which are encountered among "identical" tubes, a situation which necessitates extensive tube tests. The following list indicates some of the factors which have been shown experimentally to be capable of affecting leakage. The list should not be considered exhaustive:

1. Purity of the insulation coating: Purity of the insulation coating has been repeatedly shown to be a major factor in heater-cathode leakage. As a general rule, introduction of most contaminants increases both $+I_{HK}$ and $-I_{HK}$, although not to the same extent. While not all contaminants affect leakage seriously, certain metallic oxides such as Na_2O, Li_2O, K_2O, Fe_2O_3, ZrO_2, ThO_2, and SiO_2 do contribute, especially to $+I_{HK}$ leakage. Certain anion groups such as sulfates, chlorides, and nitrates show definite evidence of contributing to leakage, especially to high $-I_{HK}$.

2. Cathode-base-metal composition: The presence of certain additives in the cathode base metal can cause increased leakage, both initially and after a "heater negative" life. Magnesium, calcium,[19] and strontium have all been shown to increase leakage levels.

3. Heater size: It is usually possible to predict with accuracy which tube types will have continual heater-cathode leakage problems. As has been discussed previously, such tube types have large heaters with large amounts of heater wire and heater coating within the cathode sleeve.

4. Getter flash: Getter flash can deposit on the heater through the open ends of the cathode to produce leakage and hum. On types where low hum is desired, the cathode should be covered or the get-

ter should be mounted on the side of the tube.

5. Tube parts: Tests in which different tube parts have been studied have shown significant differences in heater-to-cathode-leakage levels attributable to tube parts. Certain plate materials and increased thickness of the cathode coating have been two factors which have increased leakage and hum. A high correlation is also frequently observed between high leakage and poor electrical characteristics. The amount of gas given off during tube manufacture appears to be the common source of difficulty. In a similar vein, the admission of a small amount of air to an operating low leakage tube will cause a drastic increase in heater-cathode leakage which is very difficult to age out. This may be due to the fact that once oxides have formed in the heater-cathode system, they are difficult to decompose.

6. Dirty cathodes and heaters: Contamination of cathodes and heaters has, in several instances, been identified with the source of leakage. Hydrogen firing as well as special cleaning has eliminated the complaint in many cases.

7. Insulation coating thickness: Contradictory evidence exists on the effect of insulation coating thickness. Increased coating thicknesses may cause an increase in leakage since more total contaminants are introduced into the heater-cathode system (see Tables VII and VIII). On the other hand, increased coating thickness has slightly beneficial effects in well-aged tubes.

8. Oxidized heater wires: Oxidation of the tungsten heater wire can occur in heater manufacture or tube processing. The volatile tungsten oxide will spread throughout the heater-cathode system and can contribute substantially to leakage and hum which is difficult to remove especially if the oxidation is severe. Test results have been favorable when carbon-coated heater wires (to prevent oxidation of the wire) have been tried — provided additional tube gassiness did not result.

9. Wire composition: Increased leakage and hum have been associated with heater-wire composition. Such effects are usually lowest on types using heaters with coated apices or Dowmo heater wire.

10. Coating texture: Rough heater coatings are generally reputed to be slightly superior for leakage although the difference does not appear to be important, especially in well-processed tubes.

11. Gassy tubes: Higher leakage levels are generally shown by gassy tubes.

12. Etched cathodes: Tests with some types of chemically etched or polished cathodes have resulted in slightly higher leakage levels than obtained with the untreated controls. This effect was presumably due to contaminants introduced in the chemical treatment or to the slight changes in cathode temperature due to the surface treatment.

13. Heater movement: On several occasions, tubes with initially low heater-cathode leakage have had high heater-cathode leakage when remeasured after tapping. As previously discussed, this effect is apparently due to movement of the heater within the cathode.

14. Heater type: Double-helical and folded single-helical coiled heaters have shown, as a rule, lower leakage and hum levels than folded heaters. Part of this effect is believed to be due to the high-temperature firings received by these coiled heaters in manufacture. An alternate factor could be the smaller areas of contact with the cathode walls made by the coiled heaters. An independent effect important in hum problems is reduced magnetic coupling with the electron stream when double helical or folded single-helical heaters are employed.

15. Al_2O_3 layer on the inside of cathode: Foreign tube manufacturers use a process for applying a thin (0.5-mil) coating of aluminum oxide to the inside of the cathode sleeve for tubes in which extremely low hum levels are desired. This coating will reduce the amount of material from the cathode which enters the heater coating. A somewhat similar, although apparently not quite as effective, result can be obtained by a short hot shot stage during aging in which aluminum oxide is vaporized onto the inside of the cathode sleeve.

16. Sealexing variables: Variation of sealexing conditions over a wide range has produced statistically significant, although rarely very large, differences in heater-cathode leakage. Provided the heaters are not severely oxidized or fused, much of the difference is usually taken up in aging.

17. Aging schedules: Aging schedules may have a profound effect on heater-cathode leakage as discussed in a later section.

18. Cracked heater coating: Electron emission has been reported to occur from areas of bare wire exposed through cracked coatings.[16]

19. Fused heater coatings: Heater coatings with a glassy, fused appearance have been reported (by G. R. Shaw and L. R. Shardlow) to be worse for leakage and hum.

20. Photoconductive or photo-emissive background effects are always a possibility in low level leakage measurements. Adequate precautions during measurement should be taken to ensure that they do not occur.

Heater Hum

Heater hum is related to heater-cathode leakage in that it is the result of imperfect isolation of the heater from the other components of a tube so that a small fraction of the ac power used in lighting the heater is amplified and appears as an objectionable background noise in the output of the tube. It is customary to report hum in terms of the number of microvolts which, if applied directly to the grid, would produce an equivalent disturbance.

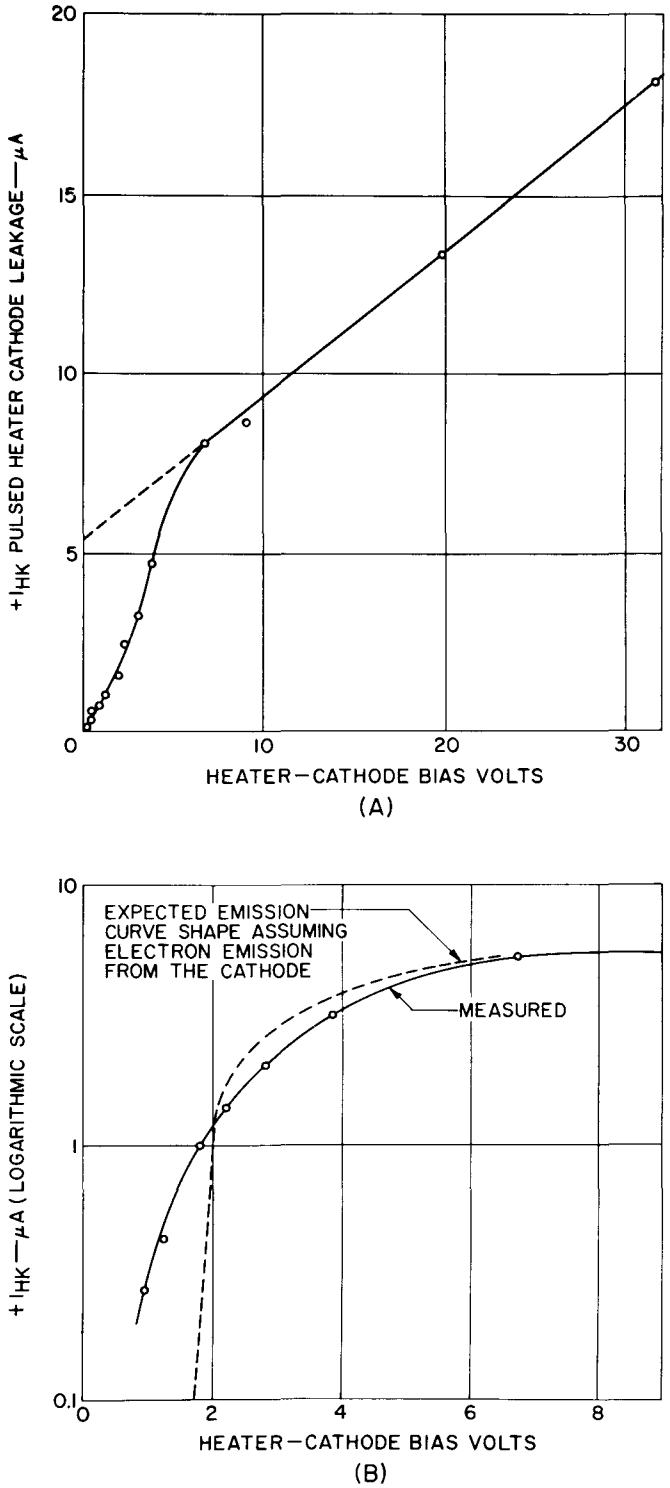


Figure 19. Typical $+I_{HK}$ Heater-Cathode Leakage Characteristics Obtained by Pulse Measurements on Heaters That Were Unipotential During Measurement (E_f = 5.5 volts)

Hum can be caused by any combination of the following factors:

1. Heater-cathode leakage
2. Heater-to-grid emission
3. Capacitative coupling between heater and grid or heater and cathode
4. Inductive coupling between heater and grid or heater and cathode
5. Modulation of the electron stream by the heater's magnetic field
6. Photoconductive or photo-emission effects caused by light striking tube parts.

Each of the above components can be evaluated separately through appropriate tests with the aid of an oscilloscope to show relationships.[20]

The total hum due to inductive and capacitive coupling between heater and cathode or grid, or both, is usually small (of the order of a microvolt) and can be remedied only by constructional changes. Modulation of the electron stream by the magnetic field of the heater is readily distinguished because the hum frequency is usually twice that of the heater-supply frequency. Such magnetic hum is mainly a function of heater design, being negligibly small for double-helical heaters, somewhat larger for single-helical heaters, and largest (up to a few microvolts) for folded heaters. It is possible that magnetic hum could be reduced by the use of a cathode material having a Curie temperature higher than normal cathode temperatures so that the magnetic fields would not penetrate it.

Leakage between heater and cathode, or heater and grid, will produce hum in circuits with large cathode or grid resistors respectively. In each case, the flow of small amounts of alternating current through the resistance will produce an ac voltage drop equivalent to one applied directly to the grid. Hum arising from leakage can be extremely variable from tube to tube.

Hum caused by electron emission from heater to grid is usually attributable to exposed lengths of bare tungsten wire on folded heaters and can usually be corrected by covering all exposed portions with insulation. Hum due to heater-cathode leakage is the usual source of most hum complaints, and the control of such hum — especially where folded heaters are used — can be a matter of considerable difficulty. A noteworthy feature of heater-cathode-leakage-type hum is that the correlation between hum and conventional leakage measurements is not high. Aside from the variation produced by the other sources of hum (e.g., magnetic hum), much of this effect arises from basic differences in the method of measurement. Conventional leakage measurements are of a static type made at heater-cathode bias of 100 volts. The type of leakage measurement to correlate best with hum, however, should be a dynamic peak-to-peak measurement made without dc bias using a center-tapped heater. In this way variations due to: (1) differences of leakage decay or polarization effects, (2) differences arising from the fact that conduction effects become a greater factor in leakage at the higher biases, and (3) differences in leakage along the length of the heater itself, would be minimized.

The elimination of hum from a tube type may call for extreme aging schedules. Due to the necessity of achieving extremely low leakages at low ac bias voltages, the effect of trace contaminants that are difficult to age

out can be unusually important. While the same general factors which produce conventional heater-cathode leakage will also produce hum, getter flash contaminations, heater wire composition, and oxidized heater wires are especially troublesome. It may also be noted that hum levels are known to be substantially affected by poor sealexing conditions or gassy tube components; substantial correlations have been found between hum and tubes which have not been properly broken down on the sealex. Folded single-helical heaters have repeatedly been shown to be substantially superior to folded heaters for hum and noise while double helical heaters are in turn superior to the folded single helical heaters. The major factors contributing to these differences are the high temperature hydrogen firing, a reduced weight of insulation (and hence, contaminants) per heater, a smaller area of contact with the cathode, the feasibility of lower design temperatures, and reduced magnetic hum.

Aging

The reduction of heater-cathode leakage and hum to commercially low levels is produced mainly during the aging of tubes. While the heater-cathode aging period can, in most instances, be fitted into a schedule required by the other tube characteristics, it may be necessary to design the aging schedule around heater-cathode aging for certain difficult tube types.

Both theory and experience indicate that the reduction of heater-cathode leakage during aging follows a reasonably straightforward exponential relationship between time and temperature. For example, essentially equivalent results from a heater-cathode viewpoint can be obtained over a fairly wide temperature range by aging tubes for a long time at the lower heater temperatures or for a relatively short time at the higher heater temperatures. Marked deviations from this general rule, however, can occur if the heater temperatures reach excessive levels with heater-cathode voltages applied since permanent damage may be done to the heater; for example, such troubles as electrolysis of cathode or heater wire materials into the insulation and fused coatings can take place.

The majority of aging schedules usually involve an initial high-temperature period of up to a few minutes duration in which 1.6 to 2.0 times the rated heater voltage is applied to the heater. For a "typical" heater this voltage might correspond to temperatures in the range of 1900 to 2025 K. Such temperature estimates may be calculated from Eq. (2a) for any given tube type. Heater-cathode voltages are most frequently omitted during this stage of aging, especially at the more severe temperature conditions. During these preliminary stages, a considerable number of leakage-producing contaminants are vaporized away from the heater-cathode system, and an insulating layer is vaporized onto the inner cathode surface. The final stage of heater-cathode aging usually consists of a longer processing at a lower temperature about 1.4 times rated heater voltage (perhaps 1800 K for a typical heater). Because the rate of diffusion and vaporization of contaminants will be substantially reduced at the lower temperature levels (see Fig. 18), it is customary, although not essential,[21] to apply ac aging voltages between heater and cathode to augment the thermal diffusion of ionic contaminants to free surfaces from which they can be vaporized out of the heater-cathode system. By this means, the speed of aging is usually substantially increased. A natural question which would arise at this point is whether the rapidly reversing polarity of an ac aging voltage would be efficient in channelling the movement of ionic contaminants. It may be noted that use of only a single dc bias voltage during aging has not generally been found feasible in the factory since, depending on the polarity of the voltage used, either $+I_{HK}$ or $-I_{HK}$ would be preferentially reduced (see Table VI). Laboratory tests using an equivalent dc heater-cathode aging voltage with polarity automatically switched every few seconds have shown evidences of somewhat increased efficiency. By systematically varying the temperature and time schedules of the heater during aging, it is usually possible to work out an aging schedule satisfactory from the viewpoint of heater-cathode leakage, hum, or other tube characteristics.

As a last resort in cases of troublesome heater-cathode leakage and hum, one or more very short and severe hot-shotting steps without heater-cathode voltages have frequently been effective in reducing leakage without necessitating prolonged aging. Such steps can cause permanent damage to a tube, however, and considerable engineering data should be amassed on any possible deleterious effects. For greatest effectiveness, and to avoid damage to the other tube parts, these hot shots should be applied to cold tubes, should last no more than 15 to 30 seconds, and should be followed by a cooling period so that only the heater will reach a high temperature. Voltages as high as 3.0 times rated heater voltage have been used[22] (near 2300 K for a typical heater) but such severe steps should be viewed with caution, since the heater coating may melt (melting point 2323 K) and the heater be permanently damaged. These high-temperature treatments invariably result in some cracking of the continuous heater coating; such cracking begins to occur when temperatures greater than 2150 K are attained. By means of such hot shots, contaminants are vaporized from all parts of the heater and a desirable thin film of insulation coating is vaporized onto the inside of the cathode. Such hot-shotting stages are generally most effective when applied near the end of the aging schedule after a good part of the original contaminants have been eliminated, because the possibilities of subsequent recontamination of the heater-cathode system are then minimized. It may be noted that this severe temperature processing can cause the heater current to increase permanently[7] by as much as 10 per cent and that the design of the heaters would probably need to be adjusted accordingly to keep the heater current centered if such a step were permanently incorporated into the aging schedule.

REFERENCES

1. Code J1
2. Code J2
3. Port, J., "Rhenium - A Promising Refractory Metal," Materials in Design Engineering, Vol. 51, No. 6, pp. 140-142, 1960
4. Wang, C. Y., Tungsten, Am. Chem. Soc. Monograph Series, Reinhold, N. Y., 1943
5. Smithells, C. J., Tungsten, A Treatise on its Met-

allurgy, Properties and Applications, Chapman and Hall, Ltd., London, 1926
6. Code J3
7. Code J4
8. Code J5
9. Code J6
10. Benjamin, M., Jour. I. E. E., 80, 419-424, London, 1937
11. Code J7
12. Navias, L., J. Amer. Ceram. Soc., 15, 248, 1932
13. Rodenhuis, K., H. Santing, H. J. M. Van Tol, Philips Technical Review, 18, 188-189, 1956
14. Metson, G. H., Electron and Electron Physics, Vol. VIII, pp. 432-439, Academic Press Inc., New York, 1956
15. Airlines Electronic Engineering Committee Report, June 26, 1956
16. Advisory Group on Electron Tubes, "Heater-Cathode Leakage", Tele-Tech. and Electronic Ind., 15, 56, 1956
17. Arizumi, T., and Tani, S., Jour. Phys. Soc. Japan, 5, 442-447, 1950
18. Code J8
19. Parkhomenko,V. S., et al, Izvest Akad Nauk S. S. S. R, Ser Fiz, 1112-1119, 1956 (Cited in Chem. Abst., 51, 9, 6321, 1957)
20. Graffunder, W., Telefunken Rohre, 12, 46, 1938
21. Code J9
22. Code J10

Heat Transfer in Receiving Tubes

O. H. Schade, Jr.

Harrison

This chapter presents methods for analyzing the heat-transfer processes that occur in receiving tubes due to conduction, convection, and radiation, and for determining the resultant electrode temperatures. The concepts used in radiant-energy transmission are described in detail because radiation is the most complicated energy transfer in a tube structure. A section concerned with the application of heat theory to tube problems is included.

HEAT TRANSFER BY STEADY, UNIDIRECTIONAL CONDUCTION

The Fourier Equation and Two Simple Solutions

The amount of heat that flows within a body depends upon the temperature gradient present, the cross-sectional area of the body, and the thermal conductivity of the body material. In 1822, Fourier proposed the equation for simple, steady heat flow by conduction:

$$q = -KA(dT/dx) \quad (1)$$

The quantity K is the thermal conductivity of the body, A is the cross-sectional area, and dT/dx is the temperature gradient. Actually, the thermal conductivity of most materials varies with the temperature of the material, but a mean value K_m over the temperature range encountered usually yields solutions of sufficient accuracy for most engineering purposes. The solution of Fourier's equation for a solid of constant cross-sectional area, as shown in Fig. 1, is:

$$q = KmA(\Delta T/x) \quad (2)$$

Another simple solution involves the flow of heat through the walls of a cylinder, as shown in Fig. 2, where the cross-sectional area is proportional to the radius of the structure. The equation for the heat flow is:

$$q = K_m A_m (\Delta T/x) \quad (3)$$

where

$$A_m = \frac{A_2 - A_1}{\ln(A_2/A_1)} \quad (4)$$

Heat flow is usually expressed in watts, length in centimeters, area in square centimeters, temperature difference in degrees Centigrade (or degrees Kelvin) and the thermal conductivity in watts/cm^2/deg C for a unit length. The preceding equation of conductive heat flow can be applied to most problems encountered in a conventional tube design.

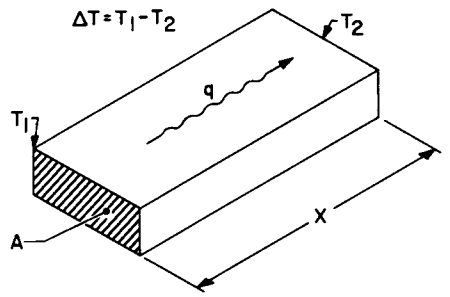


Figure 1. Dimensions of a Solid of Constant Cross-Sectional Area

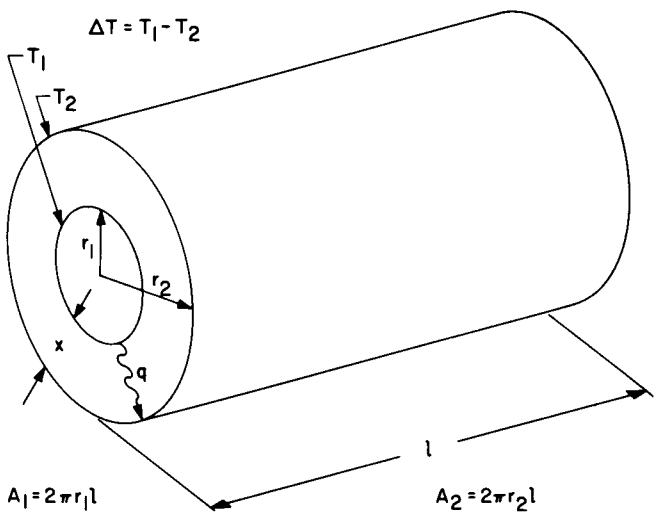


Figure 2. Dimensions of a Cylinder

The Concept of Thermal Resistance

In many cases, it is convenient to adapt the concept of thermal resistance in heat-transfer problems. Doing so permits a network of conducting bodies to be analyzed in a manner analogous to that used for electrical network analysis. The thermal resistance may be defined as the temperature difference divided by the heat flow, or

$$R = \Delta T/q \quad (5)$$

When the heat transfer occurs by steady, unidirectional

conduction, the thermal resistance may be expressed

$$R = \Delta T/q = x/K_m A = x/K_m A_m \qquad (6)$$

when Eqs. (2) and (3) are applicable. The determination of the heat flow through a sandwich of several materials of cross-sectional area A, as shown in Fig. 3, illustrates the use of thermal resistances. The total temperature drop is $(T_1 - T_4)$. The individual thermal resistances are: $R_1 = x_1/K_{m1}A$, $R_2 = x_2/K_{m2}A$, and $R_3 = x_3/K_{m3}A$. The total heat flow is $q = (T_1 - T_4)/(R_1 + R_2 + R_3) = \Sigma\Delta T/\Sigma R$.

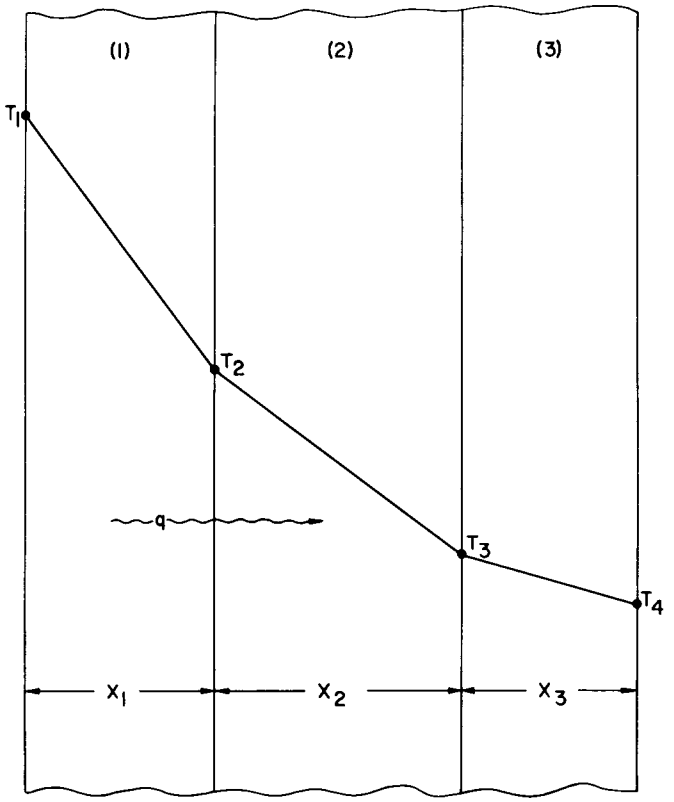


Figure 3. Heat Flow Through a Sandwich of Several Materials

The illustration is of a series arrangement of thermal resistance, but any network of series and parallel elements can be attacked by the same methods used for electrical networks. It is also possible to conceive of "convection resistances" and "radiation resistances" in the network, as will be illustrated later. The thermal resistances of various materials may be found in Figs. 15 and 16.

HEAT TRANSFER BY CONVECTION

Convection is the transfer of heat by fluid flow. Most, if not all, convection problems originate with the transfer of heat between a surface and a moving fluid. The physics of the transfer is illustrated in Fig. 4, which shows a fluid moving along the surface of a solid. Three types of flow are depicted; a laminar flow at the surface, a buffer layer of laminar flow and turbulent flow, and a region of turbulent flow at temperature T beyond the other regions. Heat is transferred through the viscous film of laminar flow by conduction and radiation. Heat is transferred through the buffer layer by conduction and convection. In the region of turbulent flow, heat is transferred primarily by convection.

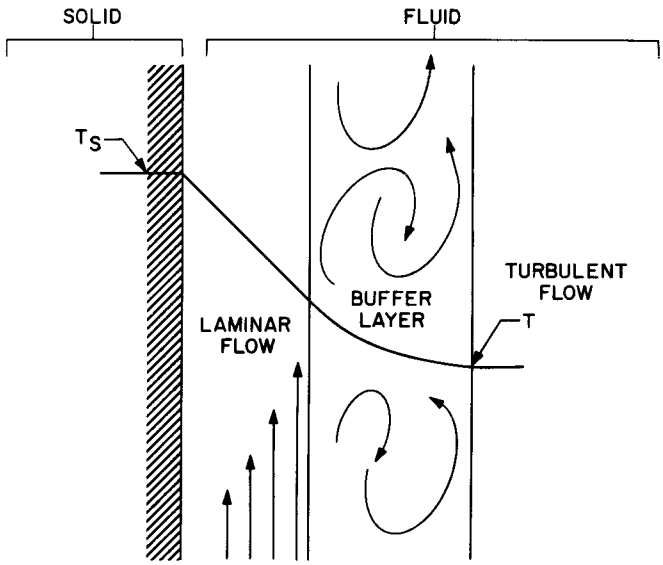


Figure 4. Fluid Flow Along the Surface of a Solid

So many variables are encountered in an attempt to solve even the simplest problem in convection heat transfer that boundary layer thickness, fluid viscosity, density, and heat capacity are assumed to remain constant. If the above-mentioned factors remain constant, the heat transfer is proportional to the area A and the temperature drop $T_S - T = \Delta T$. With the introduction of a dimensional constant h, Newton's law of cooling can be stated:

$$q = hA\,\Delta T. \qquad (7)$$

The quantity h is called the "film coefficient of heat transfer." It is evaluated by using dimensional analysis to arrive at an empirical relationship of the pertinent constants.

Many surface configurations have been evaluated, but the only problem normally encountered in tube design is the natural convection from smooth horizontal and vertical cylinders in air. The expression for h under these conditions is:[2]

$$h = 0.27\,(\Delta T/D)^{0.25} \text{ Btu/hr. ft.}^2 \text{ deg F}$$

or, in more convenient units,

$$h = 1.78\,(\Delta T/D)^{0.25} \times 10^{-4} \text{ watts/cm}^2\text{deg C} \qquad (8)$$

where D is the outside diameter of the cylinder.

THE PHENOMENON OF HEAT TRANSFER BY RADIATION THROUGH NON ABSORBING MEDIA

The Physics of Radiation

No theory that explains all the aspects of the phenomenon called "radiation" has yet been discovered.

Certain observations are best explained by the concept of radiation being an electromagnetic wave that is produced by the change in energy level of charged molecular and submolecular particles in an emitter. Because not all particles in the emitter have the same energy levels, the radiation is composed of an infinite number of frequencies (and wave lengths). The distribution of radiant energy by frequency must be similar to the distribution of energy levels among the molecules of the radiating object, and one would expect that the frequency of radiation varies somewhat like a probability curve (as will be shown later).

Radiation occurs at all temperatures and is independent of temperature difference between radiating bodies. For example, if two bodies A and B are maintained at temperatures T_a and T_b, body A will radiate a certain quantity of heat to body B, independent of the temperature T_b. Body B will radiate a certain quantity of heat to Body A independent of the temperature T_a. It is not possible to "buck-out" the radiation between bodies in the sense that (in electrical circuits) two batteries of similar voltage with like terminals attached will not be able to cause a current. It is possible that the net heat interchange between body A and body B is zero, but the force (temperature) that causes energy interchange is always present (as long as the temperature is above absolute zero). This is a statement of the "Prevost law of exchanges."

The Black-Body Concept

A body that is a perfect absorber and emitter of radiant energy is called a "black-body." The concept of a black body is useful because it provides a standard to which the absorption or emission of radiant energy by practical bodies (or surfaces) can be compared.

The characteristics of a black body can be closely approached by a model first proposed by Kirchoff: If a small opening is made in a hollow sphere whose inner surface is covered with a material that is a good absorber (such as lamp-black), radiation incident on the small opening will be almost totally absorbed within the sphere, because of the many reflections and absorptions of the energy before the much-diminished "wave" again encounters the opening.

The Distribution of Radiation

The radiation from a black body was predicted by Wien, based upon the wave theory of transmission. He proposed that the quantity of energy radiated by a black body at any particular wave length (called the "monochromatic emissive power") could be expressed

$$E_{b\lambda} = C_1 \lambda^{-5} e^{-C_2/\lambda T} \qquad (9)$$

where C_1 and C_2 are constants, λ is the wave length of the radiation, and T is the temperature of the radiating body. The dimensions of monochromatic emissive power are radiant energy at a particular wave length, per unit time, from a surface of unit area. Correlations with experimental evidence were good at short wave lengths but poor at long wave lengths. However, the wave length of maximum radiation could be determined with little error from Wien's "displacement law."

$$\lambda_{max} T = 2880 \text{ microns} \cdot \text{deg K} \qquad (10)$$

Planck introduced a corpuscular theory of radiation which premised that energy is emitted in discrete bundles, or corpuscles, rather than in a continuous form. He called these bundles of energy "quanta." The relationship found between emissive power, wave length, and temperature, as postulated by the corpuscular theory, agreed well with experimental evidence. The relationship is of the form:

$$E_{b\lambda} = C_1 \lambda^{-5} / (e^{C_2/\lambda T} - 1) \qquad (11)$$

which is similar to Wien's law. The solution for the constants yields the equation:

$$E_{b\lambda} = 3.73 \times 10^{-16} \lambda^{-5} / (e^{1.436/\lambda T} - 1) \text{ watts/cm}^2\text{micron} \qquad (12)$$

where T is expressed in degrees Kelvin and λ is in centimeters (1 micron = 10^{-4} centimeters).

The monochromatic emissive power of a black body at different temperatures is shown in Fig. 5, as plotted from Eq. (12). Notice that as the black body temperature is increased, the maxima of the curves shift to shorter wave lengths. The body must be quite hot (about 5000 K, approximately the apparent temperature of the sun) before the maxima occur near the center of the visible spectrum.

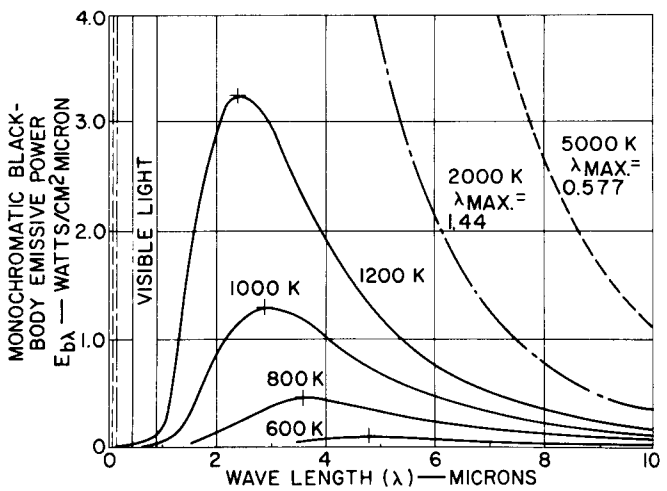


Figure 5. The Monochromatic Emissive Power of a Black Body at Various Temperatures

The total radiation at each temperature is proportional to the area lying under the curve, or by definition, $E_b = \int_0^\infty E_{b\lambda}\, d\lambda$. The surprisingly simple relationship between the total radiation emitted by a black body and its temperature and surface area was determined experimentally by Stefan and devised theoretically by Boltzmann. It is called the "Stefan-Boltzmann law":

$$q_b = \sigma A T^4 \qquad (13)$$

To obtain the heat transfer in watts, the Stefan Boltzmann constant is $\sigma = 5.67 \times 10^{-12}$ watts/cm^2(deg K)^4. The equation is probably most easily used in the form:

$$q_b = 5.67\,A(T/1000)^4 \text{ watts} \tag{14}$$

The total emissive power of a black body is

$$E_b = q_b/A = \sigma T^4 \tag{15}$$

The Gray-Body Concept

A body that emits or absorbs with less "efficiency" than a black body at the same temperature can be defined as a "gray body." The material (surface) is said to have an "emissivity" ϵ defined by the ratio

$$\epsilon = \left.\frac{E}{E_b}\right|_{T = \text{constant}}$$

At any particular wave length and temperature a body has the "spectral" emissivity

$$\epsilon_\lambda = \left.\frac{E_\lambda}{E_{b\lambda}}\right|_{T = \text{constant}} \tag{16}$$

Furthermore, the gray body has a constant spectral emissivity independent of temperature or wave length. The gray body has a radiation distribution exactly the same as that of the black body (as shown in Fig. 5), except that everywhere it is some constant fraction of the black-body radiation. The gray body has a "total" or "thermal" emissivity as defined by the ratio

$$\epsilon_t = \left.\frac{E}{E_b}\right|_{T = \text{constant}} \tag{17}$$

The gray body has a total emissive power:

$$E = \sigma \epsilon_t T^4 \tag{18}$$

Fig. 6 is the graphical solution of this equation with E versus T, and ϵ_t as a parameter.

Real materials differ from the ideal gray body in that their spectral emissivity changes with wave length and temperature and that their total emissivity changes with temperature, but for heat transfer calculations, real materials are usually assumed to be ideal gray bodies.

The Use of Spectral and Total Emissivities

Normally, there is only one instance when the tube designer need be concerned with spectral emissivity — when he determines the temperature of a body with a "disappearing filament" optical pyrometer. Pyrometers of this type are calibrated to read "true" black-body temperature. The pyrometer reads "brightness" temperatures on any body that is not black. This occurs because temperature is determined by matching light intensities; the intensity of the radiating object is matched to the intensity of the pyrometer filament. If the radiating object is not a black body, it must be at a higher temperature than a black body that emits the same light intensity. For this reason, brightness temperature must be corrected to obtain true temperature; the correction depends upon the spectral emissivity of the radiating object. Since a red filter is normally used in viewing the body, the spectral emissivity of materials is normally determined at a "standard" 0.66 microns wave length. Fig. 7 shows the correction required to obtain the true temperature of a body when its spectral emissivity is known.

Total emissivity is used whenever the heat transfer from a body is calculated. When the radiation from a body is emitted in all directions into or through an imaginary hemispherical surface, the term total "hemispherical" emissivity is used. The ratio of total emissive power of the test body to that of the black body, taken in a direction normal to the test surface, is called the total "normal" emissivity. Most information available on materials and material surfaces lists their total normal emissivity, which can be considered equal to the total hemispherical emissivity, except for well-polished surfaces; for well-polished surfaces, the hemispherical emissivity is 15 to 20 per cent greater than the normal emissivity.

SOLUTIONS OF HEAT TRANSFER EQUATION FOR RADIATION THROUGH NONABSORBING MEDIA*

Black-Body Radiation for Simple Geometries

The heat transfer equations for the black-body radiation of three simple geometries have the same solution. The geometries are:

1. Small black body radiating to large black enclosure.
2. Infinite parallel black planes.
3. Two continuous black surfaces, one enclosing the other; no negative curvature on the inner body.

For each, the net heat interchange by radiation is:

$$q_b = A_1 (T_1^4 - T_2^4), \tag{19}$$

where A_1 is the area of the hotter body.

The View Factor

When geometries become more complex than those listed above, it is convenient to adopt the idea of a "view-factor." Suppose, for example, that there are two infinite parallel black planes and plane 1, the hotter of the two, radiates heat to plane 2 as predicted by Eq. (19). If plane 2 is reduced in size, it will receive only a fraction of the heat radiated by plane 1. The heat interchange will be

$$q_b = \sigma A_1 (T_1^4 - T_2^4)\, F,$$

where F, the view factor (also called the "interchange factor," "geometrical factor," or "form factor"), represents the fraction of the heat radiated by plane 1 that is intercepted by plane 2.

Three "types" of view factor are normally considered. They are all, by definition, of the form

$$F_{1,2} = \frac{\text{heat radiated by body 1 intercepted by body 2}}{\text{total heat radiated by body 1}}.$$

*Equations and concepts primarily from Hottel, Refs. 4 and 7.

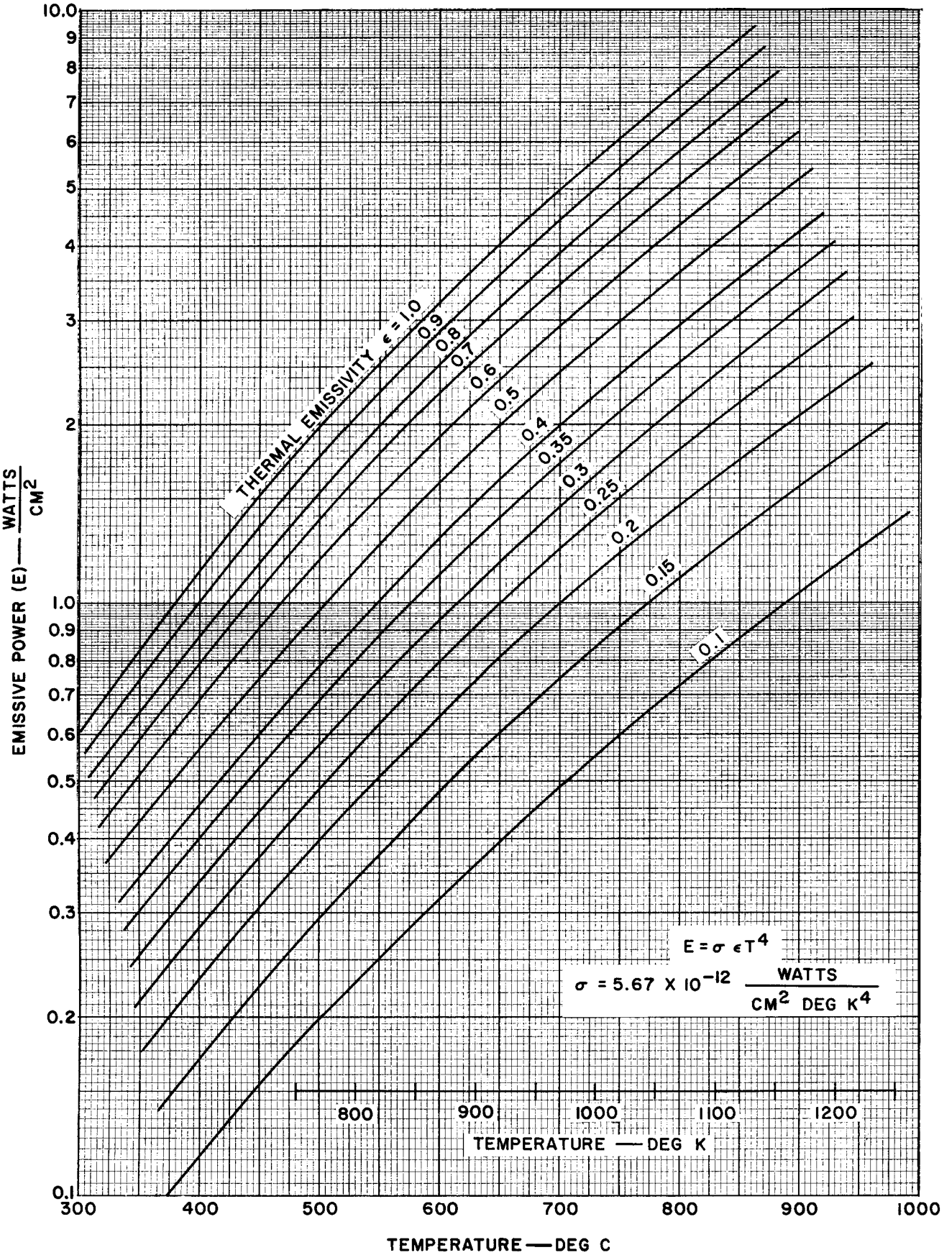


Figure 6. Emissive Power as a Function of Temperature for Various Thermal Emissivities

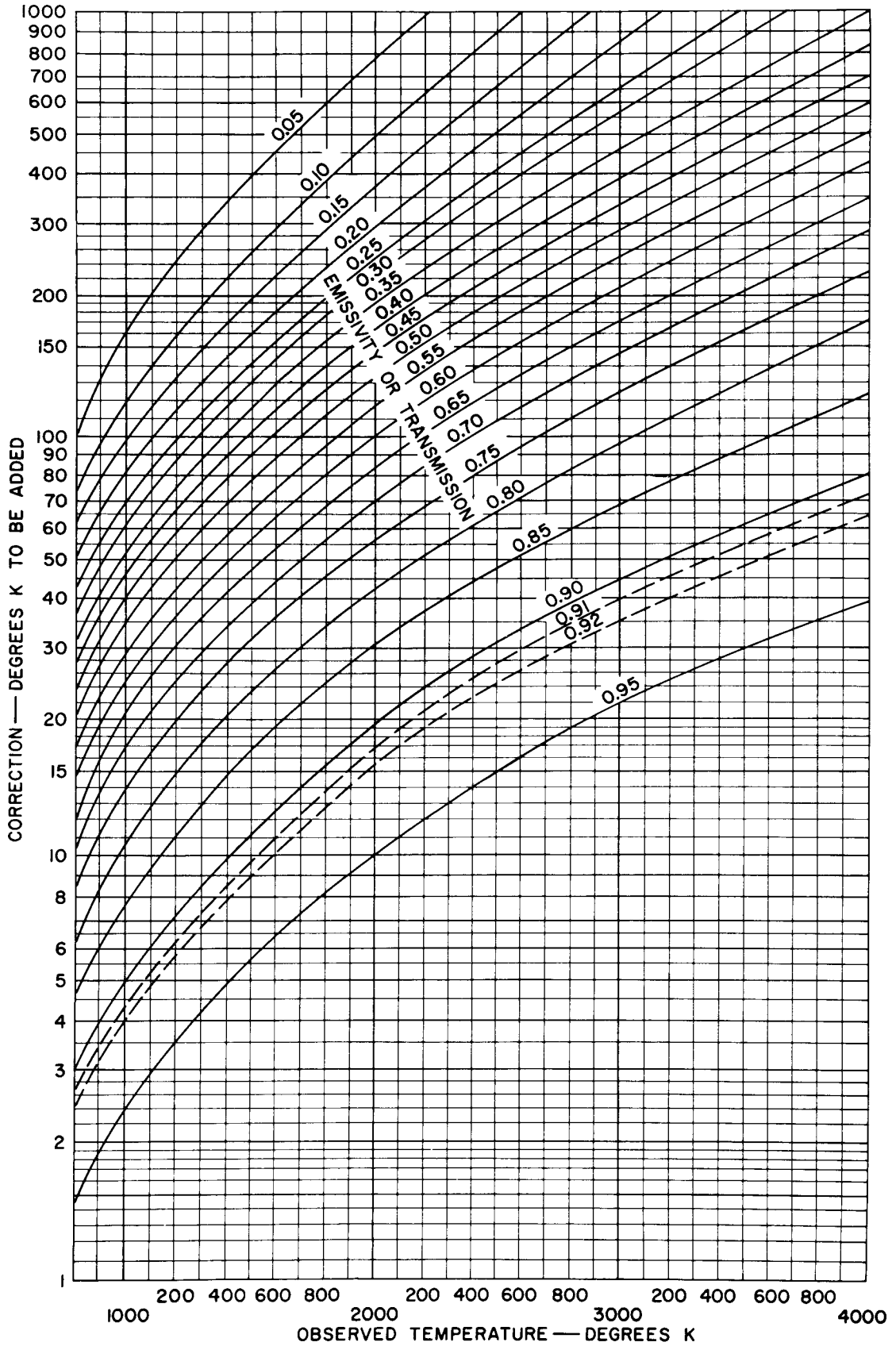


Figure 7. Corrections to Brightness Temperature Determinations by Optical Pyrometer for Transmission of Bulb and Emissivity at 0.665 Microns

The individual types are:

1. F: the view factor determined by the geometrical construction and positioning of two black bodies.
2. $\bar{F}$: the view factor, determined by geometrical considerations, which accounts for the effect of other reflecting bodies near the two black bodies of concern.
3. $\mathcal{F}$: the "over-all" view factor, which accounts for geometrical considerations, the effects of other bodies, and the emissivities of two gray bodies of concern.

The view factors may be evaluated by a mathematical approach, by graphical integration, or with the help of mechanical devices. The view factors, F and $\bar{F}$, for several geometrics are presented in Figs. 8 to 14. The factors apply to:

1. Adjacent rectangles forming angles of 60°, 90°, and 120°.
2. Long, opposed parallel rods.
3. Parallel plane structures.
4. Parallel planes and rows of tubes.

The over-all view factor $\mathcal{F}$ contains the effect of the emissivities of gray bodies. Only the simplest are accurately expressed.

Infinite Parallel Gray Planes (F = 1).

$$\mathcal{F} = \frac{1}{(1/\epsilon_1) + (1/\epsilon_2) - 1} \qquad (20)$$

A Gray Radiating Surface That is Small Compared to the Distance to Other Gray Surfaces.

$$\mathcal{F} = \epsilon_1 \epsilon_2 \cdot F \text{ (or } \bar{F}). \qquad (21)$$

A Gray Enclosed and Gray Enclosing Surface.

$$\frac{F \text{ (or } \bar{F})}{(1/\epsilon_1) + (1/\epsilon_2) - 1} < \mathcal{F} < \frac{F \text{ (or } \bar{F})}{(1/\epsilon_1) + (A_1/A_2)[(1/\epsilon_2) - 1]} \qquad (22)$$

The left-hand term is obtained assuming completely specular reflection of heat (equal angles of incidence and reflection of a wave). The right-hand term is obtained by assuming completely diffuse reflection. With practical surfaces neither extreme is correct, but the right-hand expression (diffuse reflection) is much more nearly correct for most surfaces, particularly if they are rough, oxidized or nonmetallic.

A Gray Body Enclosed by Gray Tubes.

$$\mathcal{F} \cong \frac{F \text{ (or } \bar{F})}{(1/\epsilon_1) + (A_1/A_2)[(1/\epsilon_2) - 1]} \qquad (23)$$

This is a further approximation of Eq. (22).

Pairs of Gray Surfaces Not Forming an Enclosure (General).

$$\mathcal{F} \cong \epsilon_1 \epsilon_2 \cdot F \text{ (or } \bar{F}) \qquad (24)$$

This covers long opposed parallel gray rods, adjacent gray rectangles, gray plane-parallel structures with boundaries, and the like. The factor in Eq. (24) always tends to give a somewhat low, but reasonably good result when the surfaces are designed to be good radiators or are rough, oxidized, or nonmetallic.

In general, the accuracy of calculations of radiation heat transfer in tubes will be most affected by the difficulty of predicting the emissivity of the surfaces involved. For example, most metals have a considerably lower emissivity when their surfaces are clean than when they are oxidized or covered with cathode evaporants. When the emissivities of these surfaces have been evaluated and the proper values used in the problem, the accuracy of the calculations will be truly dependent upon the approximations made in determining the heat transfer equations.

Radiation from "Composite Surfaces"

It is possible to determine the radiation from surfaces composed of the more-simple geometrical shapes for which view factors are known. The radiation from two of such "composite" surfaces will now be evaluated.

Radiation from a "Right-Angled" Surface at Constant Temperature. Fig. 17 shows a cross section of a "right-angled" surface that, for example, might be found on the half-plates of large beam-power receiving tubes. The radiation "loss" from this composite surface to the large enclosure E, which is assumed to be a black body, is to be found. Body 1 has a radiating area A_1 (that side facing the enclosure as shown), an emissivity ϵ_t, and a temperature T. Body 2 has a radiating area A_2, an emissivity ϵ_t (the same as body 1), and a temperature T; the temperature of the enclosure is T_E. Assume, for the moment, that only body 1 is hot. Assuming diffuse reflection, it radiates the quantity of heat

$q = \sigma \epsilon_t A_1 T^4$, of which

$\sigma \epsilon_t A_1 T^4 \cdot (1 - F_{1,2})$ is lost to the enclosure, and

$\sigma \epsilon_t A_1 T^4 \cdot F_{1,2}$ is intercepted by body 2. Of this intercepted heat, body 2 loses

$\sigma \epsilon_t A_1 T^4 \, F_{1,2} (1-\epsilon_t) (1 - F_{2,1})$ to the enclosure and reflects to body 1 the amount

$\sigma \epsilon_t A_1 T^4 \, F_{1,2} (1-\epsilon_t) F_{2,1}$. Of this amount, body 1 loses to the enclosure

$$\sigma \epsilon_t A_1 T^4 F_{1,2} (1-\epsilon_t) F_{2,1} (1-\epsilon_t) (1 - F_{1,2}),$$

and reflects to body 2

$$\sigma \epsilon_t A_1 T^4 F_{1,2} (1-\epsilon_t) F_{2,1} (1-\epsilon_t) F_{1,2}$$

and so forth.

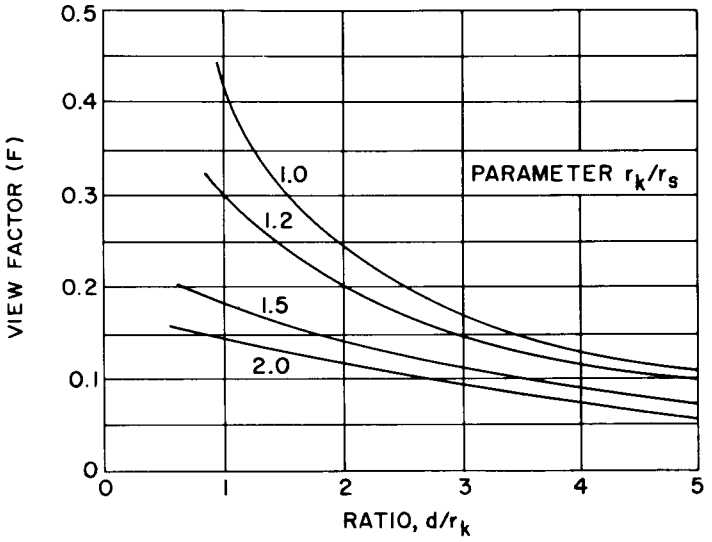
So far, the total losses to the enclosure due to the radiation from body 1 are:

$$q_1 = \sigma \epsilon_t A_1 T^4 (1 - F_{1,2}) + \sigma \epsilon_t A_1 T^4 F_{1,2} (1-\epsilon_t) (1 - F_{2,1}) +$$

$$\sigma\epsilon_t A_1 T^4 F_{1,2} F_{2,1} (1-\epsilon_t)^2 (1-F_{1,2}) \text{ or}$$

$$q_1 = \sigma\epsilon_t A_1 T^4 \left[(1-F_{1,2}) + F_{1,2}(1-\epsilon_t)(1-F_{2,1}) + F_{1,2}F_{2,1} \cdot (1-F_{1,2})(1-\epsilon_t)^2\right]$$

Examine the magnitudes of the bracketed quantities. If the structure is geometrically designed to be a radiator (to have high losses to the enclosure), and has a reasonably high emissivity, the third term is negligible. For example, if $F_{1,2} = F_{2,1} = 0.2$ and $\epsilon_t = 0.5$ (normally it



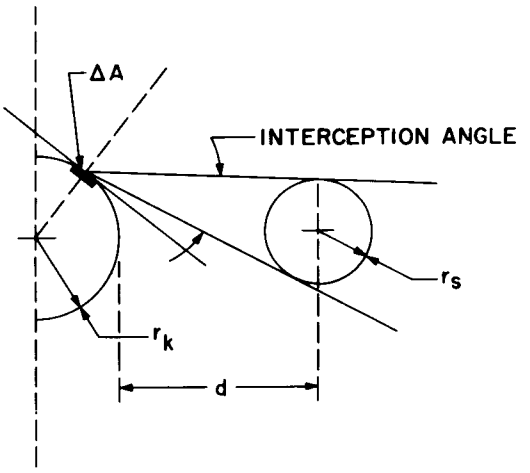


Figure 8. View Factor for Long, Opposed Parallel Rods

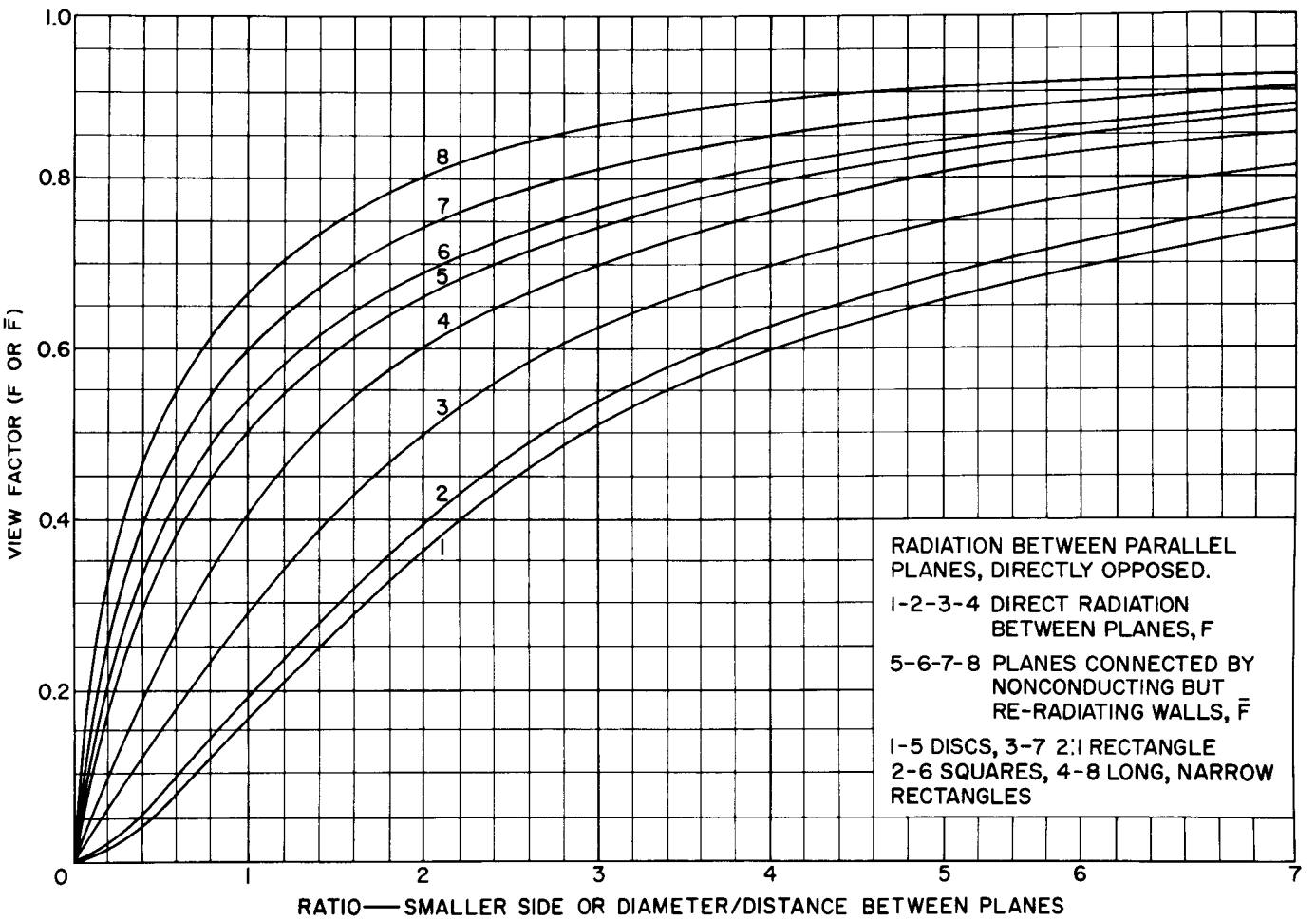


Figure 9. View Factor for Opposed Parallel Disks, Squares, and Rectangles

is even higher than this), the bracketed terms have the values

$$\left[0.8 + 0.08 + 0.008\right]$$

and the third term is surely negligible. Therefore, let the radiation loss from body 1 to the black enclosure of the quantity be defined as

$$q_1 = \sigma\epsilon_t A_1 T^4 \left[(1-F_{1,2})+F_{1,2}(1-F_{2,1})(1-\epsilon_t)\right],$$

when only body 1 is hot. Now, if only body 2 were hot, it would have a loss of the same form, or

$$q_2 = \sigma\epsilon_t A_2 T^4 \left[(1-F_{2,1})+F_{2,1}(1-F_{1,2})(1-\epsilon_t)\right],$$

which can be obtained by a similar method of attack. The total loss, with both body 1 and body 2 hot, can be factored to the form

$$q = q_1 + q_2 = \sigma\epsilon_t T^4 \left\{(A_1 + A_2)\left[1 + F_{1,2}F_{2,1}(\epsilon_t - 1)\right] - \epsilon_t (A_1 F_{1,2} + A_2 F_{2,1})\right\}$$

When the large black enclosure is also hot, the net heat transfer to the enclosure is

$$q = \sigma\epsilon_t(T^4 - T_E^4)\left\{(A_1 + A_2)\left[1 + F_{1,2}F_{2,1}(\epsilon_t - 1)\right] - \epsilon_t (A_1 F_{1,2} + A_2 F_{2,1})\right\} \quad (25)$$

If the radiating surfaces are black, $\epsilon_t = 1$ and

$$q_b = \sigma(T^4 - T_E^4)\left[A_1(1 - F_{1,2}) + A_2(1 - F_{2,1})\right], \quad (26)$$

indicating, as one would expect, that there are no reflections between the radiating surfaces. This expression may be further simplified by recognizing that there can be no net heat interchange between black bodies at the same temperature, independent of their geometries, so that

$$A_1 F_{1,2} = A_2 F_{2,1} .$$

This equation further simplifies the black "composite" surface radiation to

$$q_b = \sigma(T^4 - T_E^4)\left[A_1(1 - 2F_{1,2}) + A_2\right], \quad (27)$$

where it is necessary to determine only one view factor. When Eqs. (25), (26) or (27) are applied to a practical structure such as a plate, an average plate temperature and enclosure temperature should be used. The view factor may be obtained from Fig. 12

A U-Shaped Radiating Surface at Constant Temperature. Fig. 18 shows a U-shaped surface radiating to a large black enclosure. To find the radiation loss from within the U, a method may be used similar to that used for the right-angled surface. The total loss from each body, including reflection losses, is calculated assuming that only one body at a time is hot. It is necessary to consult two view factor figures to use any equations that might be derived. For example, the view factor $F_{1,3}$ is determined by the heat interchange in plane-parallel structures, while the view factor $F_{1,2}$ is determined by the heat interchange between adjacent, perpendicular planes.

Assuming diffuse reflection with only body 1 hot, body 1 radiates

$\sigma\epsilon_t A_1 T^4 (1-F_{1,2}-F_{1,3})$ to the enclosure,

$\sigma\epsilon_t A_1 T^4 (F_{1,2})$ to body 2 and

$\sigma\epsilon_t A_1 T^4 (F_{1,3})$ to body 3.

The radiation distribution at body 2, due to the heat from body 1, is

$\sigma\epsilon_t A_1 T^4 F_{1,2}(1-\epsilon_t) F_{2,1}$ reflected to body 1,

$\sigma\epsilon_t A_1 T^4 F_{1,2}(1-\epsilon_t) F_{2,3}$ reflected to body 3, and

$\sigma\epsilon_t A_1 T^4 F_{1,2}(1-\epsilon_t)(1-F_{2,1}-F_{2,3})$ lost to the enclosure.

The radiation distribution at body 3 due to the heat from body 1 is

$\sigma\epsilon_t A_1 T^4 F_{1,3}(1-\epsilon_t) F_{3,1}$ reflected to body 1,

$\sigma\epsilon_t A_1 T^4 F_{1,3}(1-\epsilon_t) F_{3,2}$ reflected to body 2, and

$\sigma\epsilon_t A_1 T^4 F_{1,3}(1-\epsilon_t)(1-F_{3,1}-F_{3,2})$ lost to the enclosure.

Further tracing of the heat reflections leads to loss components bearing $(1-\epsilon_t)^2$ terms, which may be neglected if the same assumptions are made as in the preceding case. The total radiation loss to the enclosure due to body 1 is then

$$q_1 = \sigma\epsilon_t A_1 T^4 (1-F_{1,2}-F_{1,3}) + \sigma\epsilon_t A_1 T^4 F_{1,2}(1-\epsilon_t)(1-F_{2,1}-F_{2,3}) + \sigma\epsilon_t A_1 T^4 F_{1,3}(1-\epsilon_t)(1-F_{3,1}-F_{3,2})$$

Since body 3 is geometrically similar to body 1, it has the same heat loss to the enclosure. The heat loss from body 2 is evaluated in similar fashion. The total loss to the enclosure is

$$q = q_1 + q_2 + q_3 = 2q_1 + q_2.$$

By geometry, certain view factors have the same value:

$$F_{2,1} = F_{2,3}$$

$$F_{1,2} = F_{3,2}$$

$$F_{1,3} = F_{3,1}$$

By summing, simplifying, factoring, and accounting for the temperature of the large black enclosure, it is found that the net heat interchange is

$$q = 2\sigma\epsilon_t(T^4-T_E^4)\Big\{(1-F_{1,2}-F_{1,3})\Big[A_1+(1-\epsilon_t)(A_1F_{1,3}+A_2F_{2,1})+(1-2F_{2,1})\Big[(A_2/2)+(1-\epsilon_t)A_1F_{1,2}\Big]\Big\}. \quad (28)$$

If the radiating surface is black, $\epsilon_t = 1$ and Eq. (28) reduces to

$$q_b = \sigma(T^4-T_E^4)\Big[2A_1(1-F_{1,2}-F_{1,3})+A_2(1-2F_{2,1})\Big], \quad (29)$$

indicating that there are two surfaces of area A_1 losing the fractional heat $(1-F_{1,2}-F_{1,3})$ to the enclosure, and a surface A_2 losing the fractional heat $(1-2F_{2,1})$ to the enclosure. As for the right-angled surface, the equality of heat interchange between black bodies at the same temperatures requires that

$$A_1F_{1,2} = A_2F_{2,1}$$

This equality may be applied to Eq. (29) to further simplify it, if desired.

THE APPLICATION OF THEORY TO HEAT TRANSFER PROBLEMS IN TUBE STRUCTURES

Heat Transfer Circuits of the Diode

It is often convenient to think of electrical circuits,

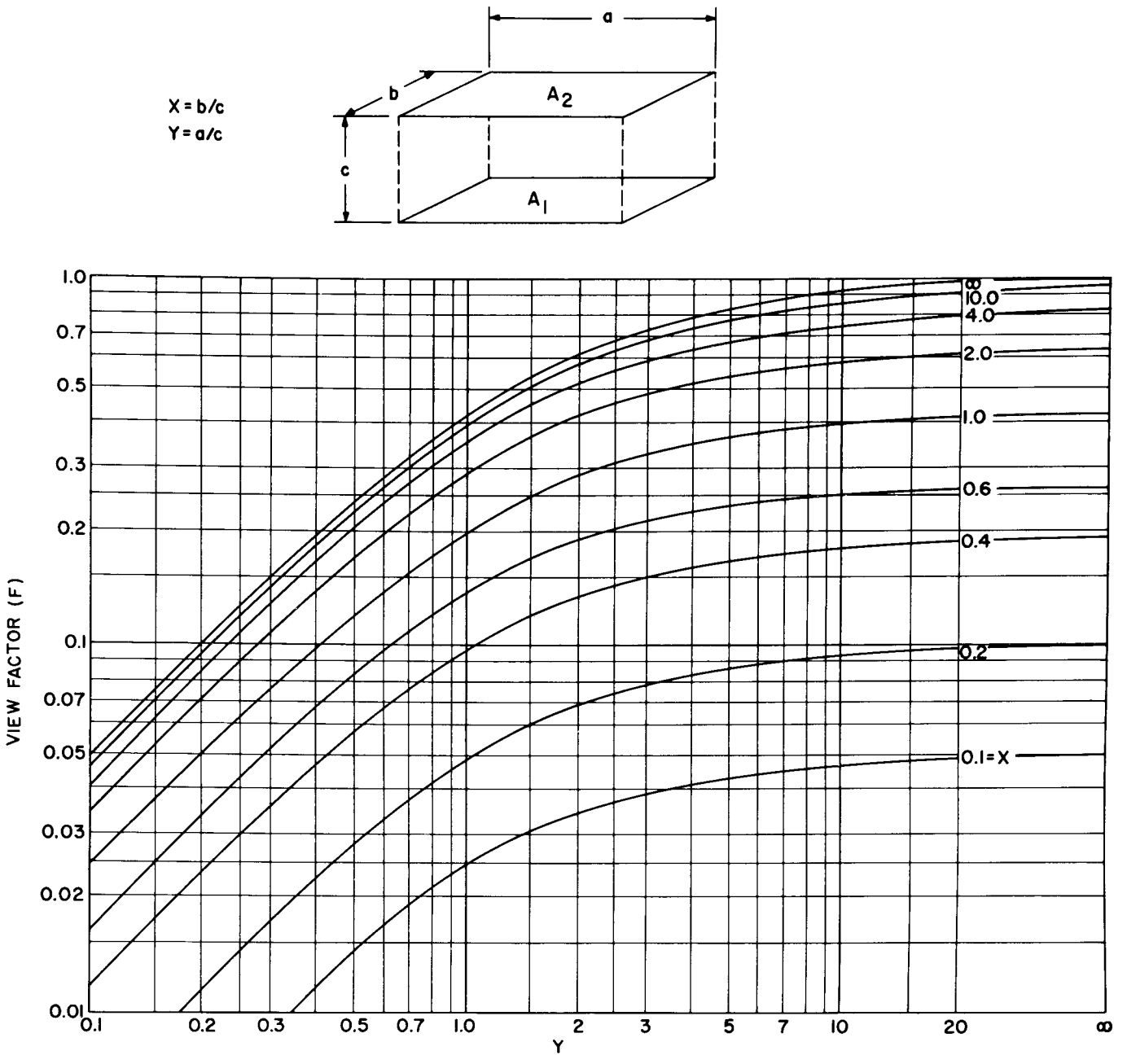


Figure 10. View Factor for Directly Opposed Parallel Planes

or networks, that would produce current flows similar to the heat flows encountered in tubes. It then becomes a simple matter to understand the types of heat flow which might be encountered in a tube problem and to appreciate their effects on electrode temperatures resulting from the magnitudes and directions of the heat flows. Fig. 19 shows a heat-transfer circuit of a diode. The heater is at "potential" T_H with respect to the ambient T_A, and forces heat to flow through the various resistances shown. The temperatures of the cathode, plate, and enclosure are designated T_k, T_p, and T_E, respectively. It is assumed that the enclosure is at a constant temperature for the purposes of this example, and that there are no temperature gradients on the various electrodes. Heat flow through any insulating spacers (micas) has been neglected. The resistance r_0 is not of concern because the heater temperature and power input are usually known or easily calculated. The total heater power divides among the resistances r_1, r_2, r_3, and r_4. The resistance r_1 is the thermal "conduction" resistance of the heater legs, which in most tubes is high enough so that any heat flow through it is relatively small. The heat flow through resistance r_2 is the radiation "end loss" from the heater at the ends of the cathode. Heat is transferred between heater and cathode through resistances r_3 and r_4, r_3 representing the conduction through the heater coating (normally a very small amount) and r_4 the radiation transfer. At the cathode, heat is transferred by conduction through a cathode tab, r_5 (normally very small), and radiation to the plate through r_6. The plate loses heat to the enclosure by stem-lead conduction through r_7 (which in "high-dissipation" structures can usually be neglected) and radiation through r_8. The enclosure loses heat to the surrounding medium by conduction, convection, and radiation through r_9, r_{10}, and r_{11}. The quantities normally of interest are the temperatures T_H, T_k, T_p, and T_E. Temperature T_k can be measured accurately with an optical pyrometer; T_E can be measured with a thermocouple or "Tempilac," or it can be estimated by using Fig. 34. T_H is quickly approximated (roughly) by using the Kauzmann nomograph for heater design or, more accurately, by referring to date on the resistivity of tungsten at different temperatures. To determine the temperature T_p, with T_k and T_E known, the simplified circuit of Fig. 20 may be used. The normal procedure is to assume an average plate temperature and to calculate the net heat transfer between cathode and plate and between the plate and enclosure. These two heat transfers will be equal when the proper assumption for T_p has been made. If an electrical dissipation appears at the plate, the two heat transfers will not, of course, be equal. In this case, the plate will assume a tem-

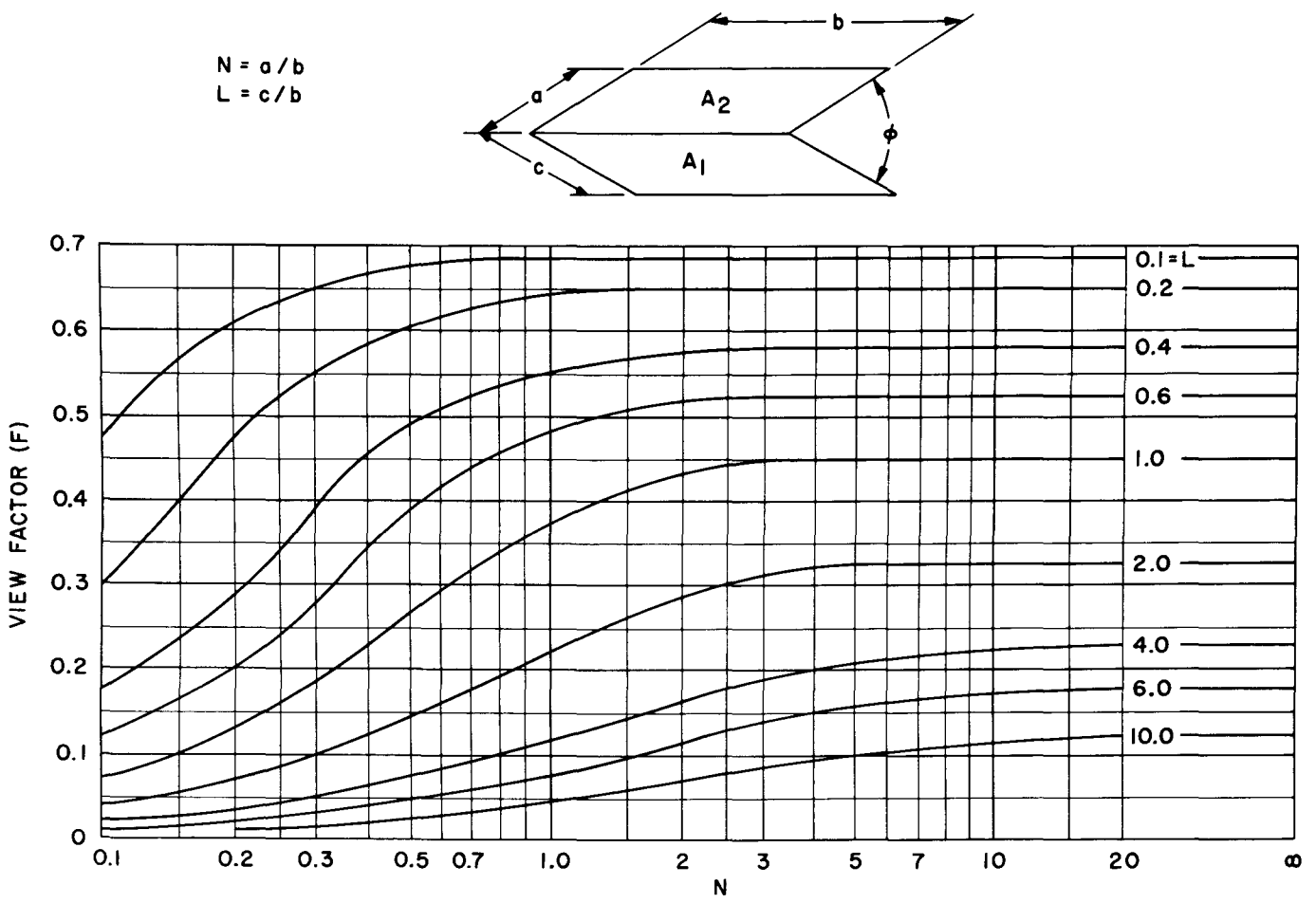


Figure 11. View Factor for Direct Radiation Between Adjacent Rectangles ($\phi = 60^o$)

perature such that it loses the amount that it intercepts from the cathode, plus the electrical dissipation. The equations for radiation heat transfer between cathode and plate and plate and enclosure are of the form

$$q = \sigma A_1 (T_1^4 - T_2^4) \mathcal{F},$$

where $\mathcal{F}$ is normally described by Eq. (22). The equation for cathode-to-plate radiation is:

$$q = \sigma A_k (T_k^4 - T_p^4) \frac{F}{(1/\epsilon_k) + (A_k/A_p)\left[(1/\epsilon_p)-1\right]} \quad (30)$$

where F = 1 if the cathode is completely surrounded by the plate (the usual approximation); for plate to enclosure radiation, the equation is:

$$q = \sigma A_p (T_p^4 - T_E^4) \frac{F}{(1/\epsilon_p) + (A_p/A_E)\left[(1/\epsilon_E)-1\right]} \quad (31)$$

or, if the enclosure is glass, where the assumption that $\epsilon_E = 1$ may be made,

$$q = \sigma A_p (T_p^4 - T_E^4) \epsilon_p F \quad (32)$$

If the plate surface is geometrically reasonably smooth, that is, if it has only small portions of its surface capable of "viewing" other portions, F in Eq. (32) is unity.

To avoid complications, it is usually convenient to use the coated area of the cathode in radiation calculations. This assumption does not normally produce serious errors in calculations requiring the usual engineering accuracy. If the plate is assumed to be considerably hotter than the enclosure, such that

$$T_p^4 >> T_E^4 ,$$

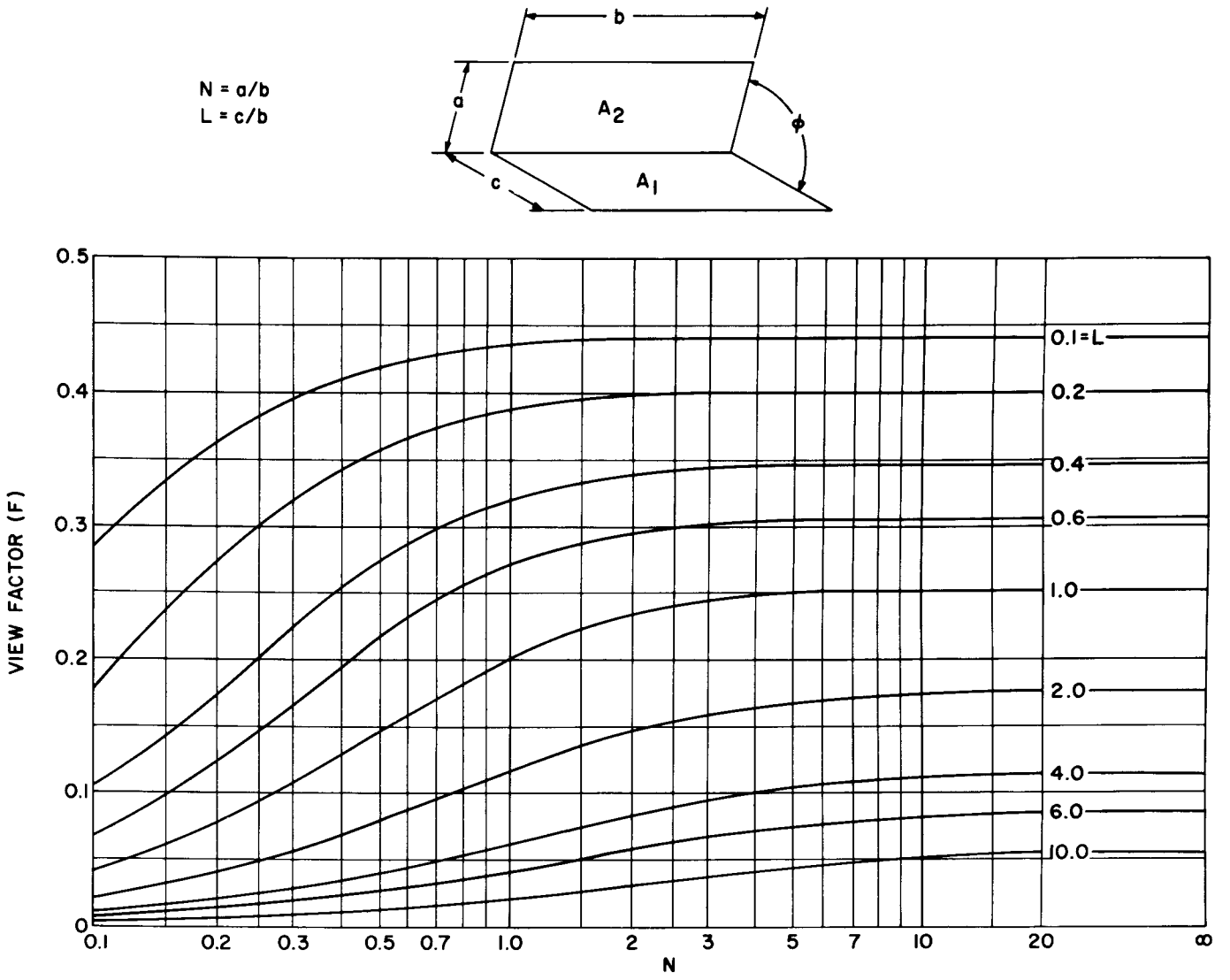


Figure 12. View Factor for Direct Radiation Between Adjacent Rectangles (Ø = 90°)

then the radiation from a relatively "smooth" plate can be quickly approximated by reference to Fig. 6. The heat conducted from the plate by a stem lead will be discussed later.

Heat Transfer Circuits of the Triode

The heat transfer circuit of the triode is shown in Fig. 21. The temperatures T_k, T_s, and T_E are assumed to be known, so that the simplified circuit may be used. Heat is radiated from the cathode to the plate through r_1 and from the cathode to the grid through r_2. The grid loses heat by conduction through r_3 to the stem at temperature T_s. It may either lose or gain heat through r_4, depending upon whether T_p is higher or lower than T_G. As in the case of the diode, the plate loses heat to the enclosure by radiation and conduction through r_5 and r_6 and, similar to the grid, may either lose or gain heat by radiation through r_4. To avoid an over-complicated trial-and-error solution for the average plate and grid temperatures, an approximate circuit analysis will be presented. It will be assumed that all the radiation incident on the plate is due to the heat transfer through r_1. The total heat the plate must lose through r_5 and r_6 to T_E is then composed of the radiation through r_1, and the plate dissipation. The average plate temperature may then be solved for by the same method used in the diode analysis. The radiation from cathode to plate differs, however, from that of the diode because the grid intercepts a fraction of it, F_{k1} (as may be determined from Fig. 14). The expression for the cathode-to-plate radiation for the triode is

$$q = \sigma A_k (T_k^4 - T_p^4) \frac{F \cdot (1 - F_{k1})}{(1/\epsilon_k) + (A_k/A_p)\left[(1/\epsilon_p) - 1\right]} \qquad (33)$$

where the usual approximation, again, is that $F = 1$.

Once the average plate temperature T_p is established, the average grid temperature can be calculated. The simplified heat transfer circuit is shown in Fig. 22. The procedure again involves a trial-and-error solution, assuming different values of T_g until the radiated heat through r_2 is equal to the algebraic sum of the conducted and radiated heat through r_3 and r_4. The

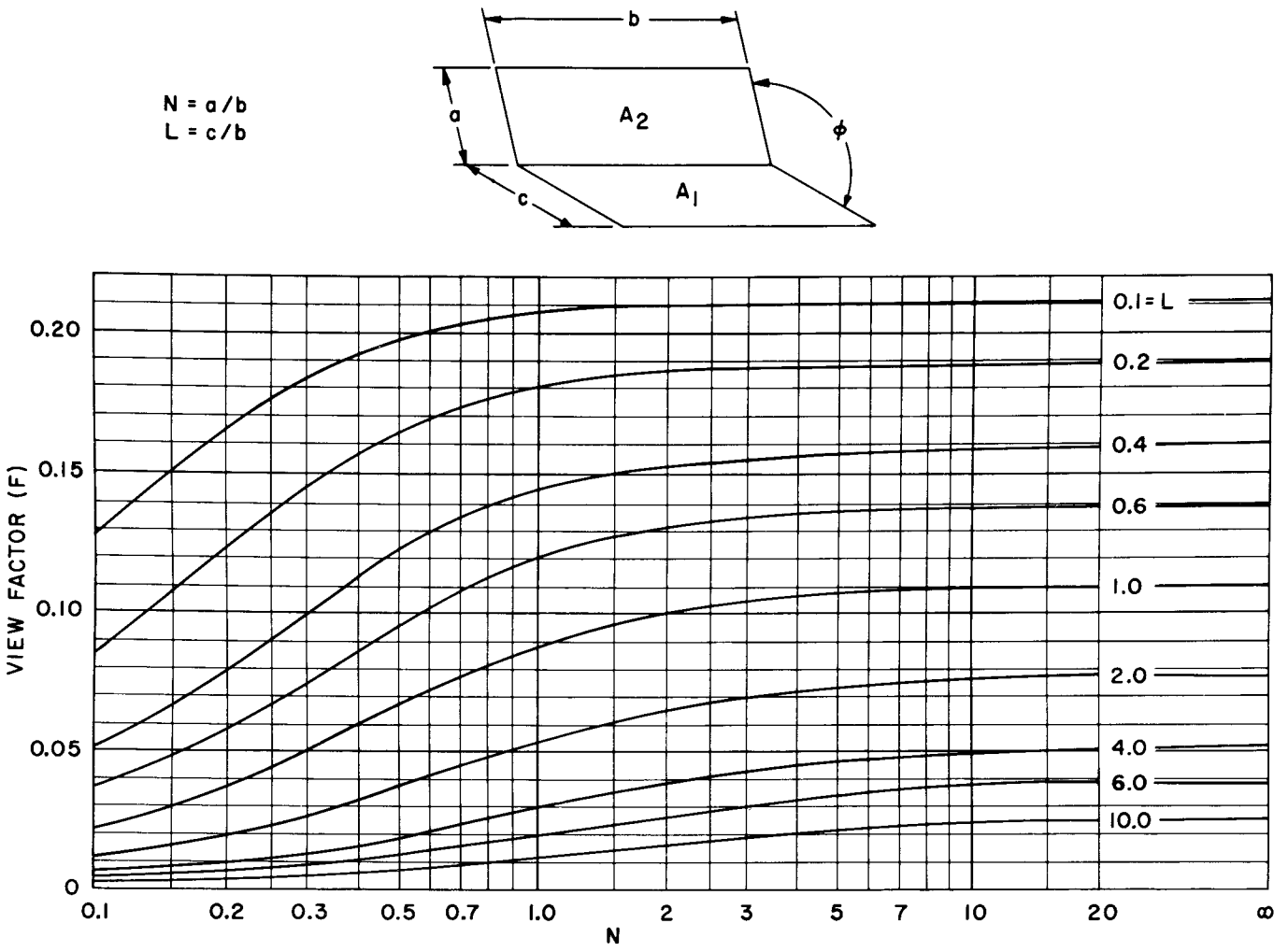


Figure 13. View Factor for Direct Radiation Between Adjacent Rectangles ($\phi = 120^o$)

cathode-to-grid radiation through r_2 can be expressed

$$q = \sigma A_k (T_k^4 - T_g^4) \frac{F_{k1}}{(1/\epsilon_k) + (A_k/A_g)\left[(1/\epsilon_g) - 1\right]} \quad (34)$$

where A_g is the area of a solid grid plane of dimensions "finished-length" times the "length of a turn of wire". If the grid is hotter than the plate, the grid-to-plate radiation will be

$$q = \sigma A_g (T_g^4 - T_p^4) \frac{F_{1p}}{(1/\epsilon_g) + (A_g/A_p)\left[(1/\epsilon_p) - 1\right]} \quad (35)$$

where $F_{1p} = F_{k1}$. If the plate is hotter than the grid, the radiation interchange is

$$q = \sigma A_p (T_p^4 - T_g^4) \frac{F_{p1}}{(1/\epsilon_p) + (A_p/A_g)\left[(1/\epsilon_g) - 1\right]} \quad (36)$$

where $F_{p1} = F_{1p} = F_{k1}$. At this point, it is necessary that the heat conducted from the grid be evaluated to complete the analysis.

Heat Transfer by Conduction Through the Grid and Stem Leads

It is possible to determine the temperature gradient or the grid, and an average grid temperature due to conduction heat flow by considering the grid to be a network of thermal resistances. Two cases will be treated: one in which a stem lead is welded to each siderod and heat flow is symmetrical about the grid center line, and the other in which there is a stem lead welded to only one siderod.

Stem Lead to Each Siderod. Fig. 23 shows the thermal resistances in one side of a grid structure which has been divided into five groups of lateral wires. Each group of lateral wires is composed of three resistances r on one side of the center line, about which fractional

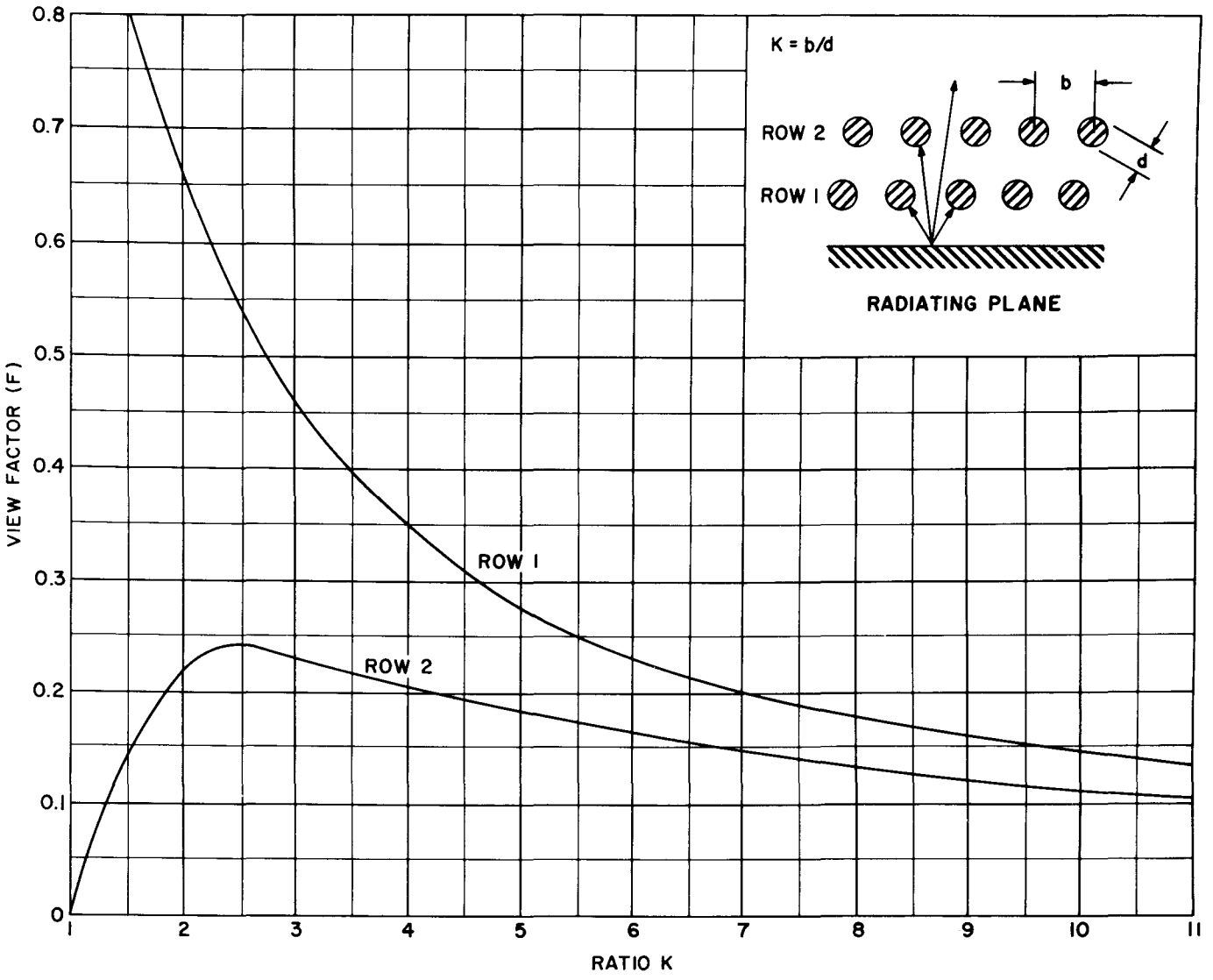


Figure 14. View Factor for Rows of Long Tubes

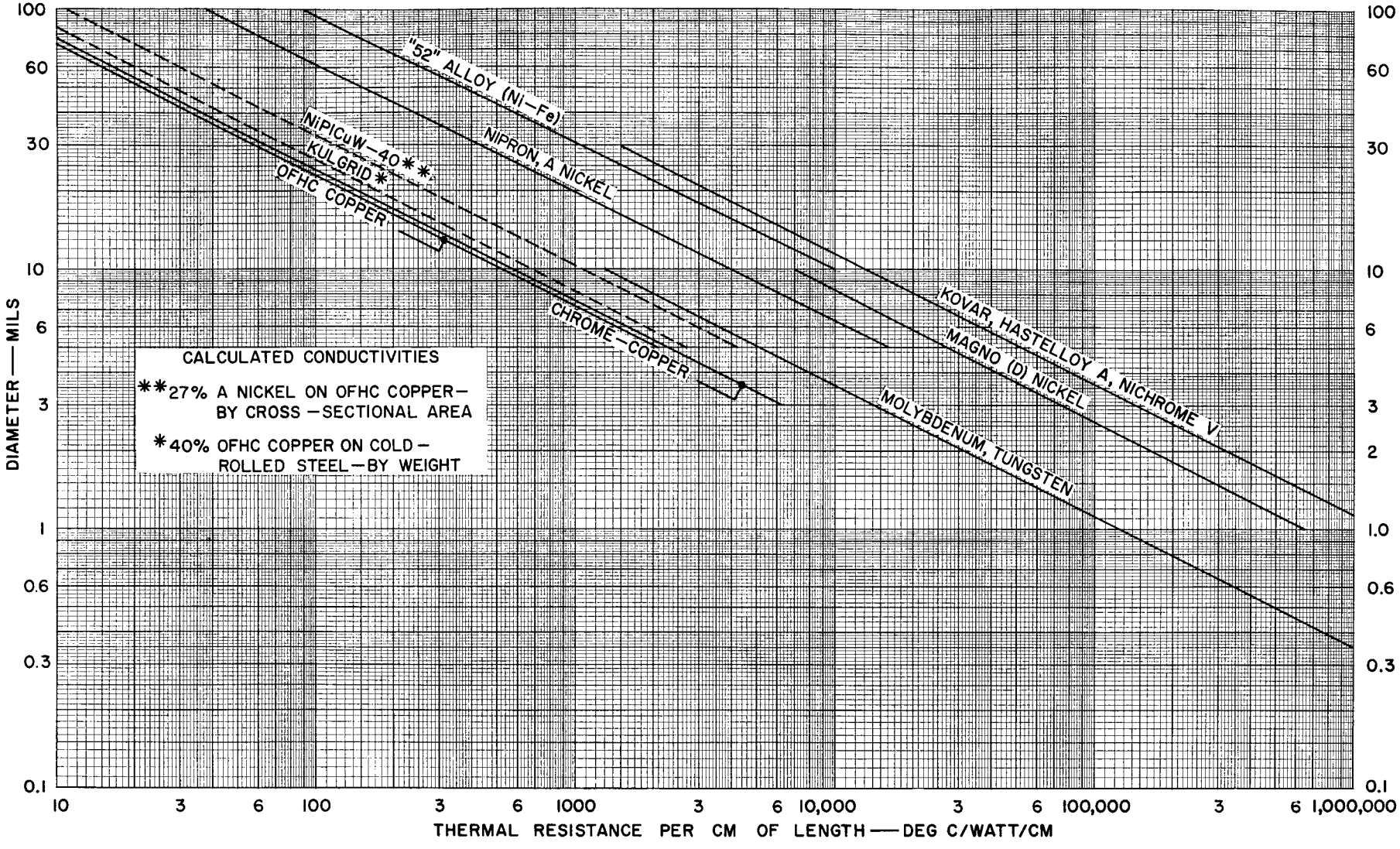


Figure 15. Thermal Resistance of Round Wires (300 C)

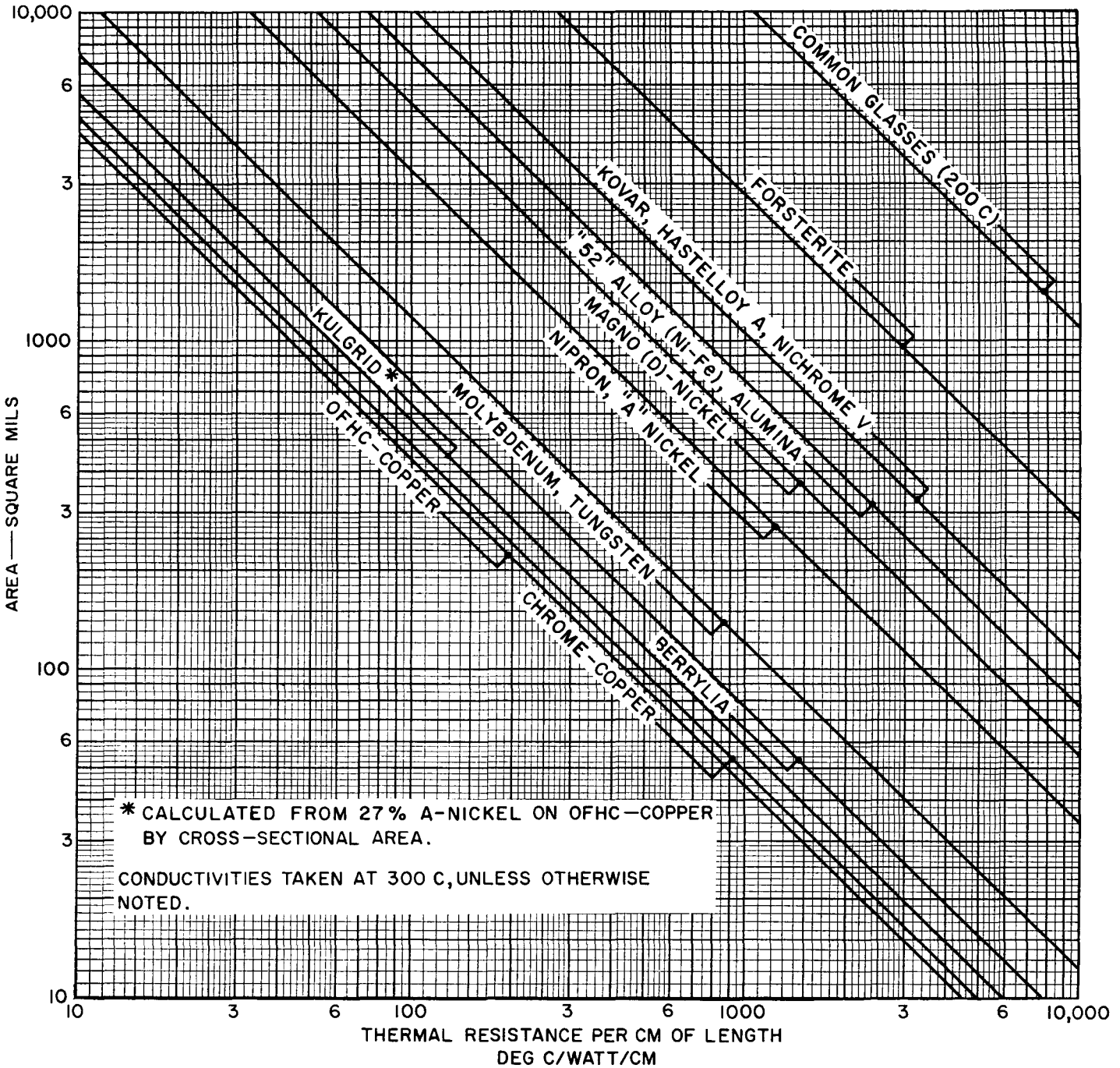


Figure 16. Thermal Resistance of Some Tube Materials

heat inputs q are assumed to be evenly distributed over the grid surface. These fractional heat inputs must flow through the lateral-wire-group resistances to the siderods, where they join the heat from the wire groups on the other side of the siderods and flow down through the siderod and stem lead to the glass part of the stem at temperature T_S. If T_S is assumed to be known, the temperature rises from the glass to the wire groups, as well as the temperature rise along the wire groups, can be calculated.

The temperature rise along the stem lead is

$$(2.15\ q)\ R_L,$$

so that

$$T_L = T_S + 30\ q\ R_L,$$

where R_L is the thermal resistance of a stem lead. There is an additional temperature rise from the stem-lead-to-siderod junction at T_L to the bottom of the first lateral wire group of the quantity

$$30\ q\ R_x,$$

where R_x is the siderod resistance from the junction to the bottom of the lateral wire. The additional rise to the average siderod temperature T_{sr} is

$$\frac{R_{f_L}}{10} \cdot 30q + \frac{R_{f_L}}{5} \cdot 24q + \frac{R_{f_L}}{5} \cdot 18q,$$

where R_{f_L} is the resistance of the "finished length" of the grid siderod. The temperature T_{sr} may then be expressed,

$$T_{sr} = T_L + 30qR_x + 30q\frac{R_{f_L}}{10} + 24q\frac{R_{f_L}}{5} + 18q\frac{R_{f_L}}{5}.$$

The temperature rise through the third lateral wire group to the center line of the grid is

$$3qr + 2qr + qr = 6qr.$$

If the average grid temperature is considered to be the average temperature of this group of wires,

$$T_g = T_{sr} + 3qr.$$

By definition, the total heat to be conducted from the grid is

$$qc = 60q$$

The geometry of the network chosen yields the relationship between the fractional lateral wire resistances r and the resistance of one finished turn of wire R_t,

$$r = (1/12)\,R_t \cdot (5/N)$$

where N is the number of finished turns on the grid. These quantities may be substituted into the expressions for the average siderod temperature and the average grid temperature, yielding

$$T_{sr} = T_L + qc\left[(R_x/2) + (19/100)\,R_{fL}\right] \quad (37)$$

and

$$T_g = T_L + qc\left[(R_x/2) + (19/100)\,R_{fL} + (R_t/48N)\right] \quad (38)$$

Notice that qc is the total heat to be conducted from the grid, which is not necessarily the total heat intercepted by the grid. The temperature T_L can be obtained by measuring (or estimating) T_S and adding the temperature rise along the stem lead, or by referring to stem heat-flow characteristics such as shown in Figs. 24 to 32. These characteristics indicate the temperature of a "reference point" on the stem lead a certain distance from the fillet and at the inner fillet, as a function of the heat flowing through the lead. In this case, the quantity qc/2 flows through one stem lead. Linear interpolations to find temperatures at intermediate points on the lead are permissible.

Stem Lead to Only One Siderod. A network of resistances similar to that depicted for two siderods may be constructed. Assuming the "floating" siderod to be thermally absent from the structure, all the heat to be conducted must pass through the entire length of the lateral wires and down one siderod. The resultant temperatures are:

$$T_{sr} = T_L + qc\left[R_x + (38/100)\,R_{fL}\right] \quad (39)$$

$$T_g = T_L + qc\left[R_x + (38/100)\,R_{fL} + (R_t/16N)\right] \quad (40)$$

In this case, all the heat qc flows through a stem lead.

The preceding material provides the solutions for the heat conducted from a grid. These solutions are necessary for the average grid temperature calculations mentioned in the paragraph on heat transfer circuits of the triode.

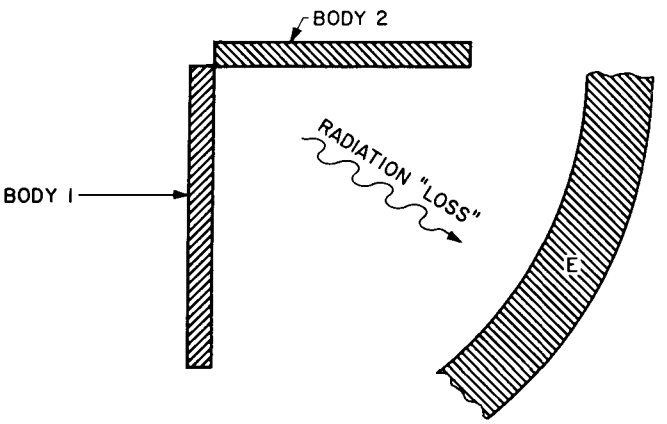


Figure 17. Cross Section of a "Right-Angled" Surface

Heat-Transfer Considerations for Multigrid Tubes

The complete solutions for heat transfer problems in multigrid tubes would be extremely tedious without using some simplifying assumptions. As in the case of the triode, the cathode-to-plate radiation can be assumed to be the only incident radiation on the plate. The equation for this transfer assumes the form

$$q = \sigma A_k(T_k^4 - T_p^4)\frac{(1 - F_{k1} - F_{k2} - F_{k3} \cdots - F_{kn})}{(1/\epsilon_k) + (A_k/A_p)\left[(1/\epsilon_p) - 1\right]} \quad (41)$$

Usually, the accuracy of calculations is not seriously impaired by neglecting radiation interchange between grids. In beam tetrodes and pentodes, the effects of beam-confining electrodes and suppressor grids can usually be neglected when calculating heat interchanges between control grid and screen grid and the plate, since their view factors are normally small. It is important to remember that any electrical dissipation on a grid must be added to the intercepted radiation to obtain the total heat that must leave the grid. If one grid radiates through another to the plate, the equation is of the form

$$q = \sigma A_{g1}(T_{g1}^4 - T_p^4)\frac{F_{1p}\,(1 - F_{2p})}{(1/\epsilon_{g1}) + (A_{g1}/A_p)\left[(1/\epsilon_p) - 1\right]} \quad (42)$$

since the intermediate grid intercepts radiation (which can be neglected in determining its temperature), and where $F_{1p} = F_{k1}$ and $F_{2p} = F_{k2}$. If the plate radiates to one grid through another, the equation is of the form

$$q = \sigma A_p(T_p^4 - T_{g1}^4)\frac{F_{p1}\,(1 - F_{p2})}{(1/\epsilon_p) + (A_p/A_{g1})\left[(1/\epsilon_{g1}) - 1\right]} \quad (43)$$

where $F_{p1} = F_{k1}$, and $F_{p2} = F_{k2}$.

While it may seem that the accuracy of the calculations is questionable in the case of multigrid tubes, the author's experience has been that the assumptions made do not seriously affect the results. For example, the grid and plate temperatures in the 6L6G were calculated for the operating conditions used in a test made by Power and Wittenberg.[8] The actual temperatures were measured by thermocouples. The results were:

	Measured	Calculated
T_p	415 C	425 C
T_{sr1}	292 C	300-330 C
T_{sr2}	515 C	525 C

The temperature of the control-grid siderod was approximated because of the effect of a grid radiator. It was assumed that the radiator operated at the average siderod temperature. The accuracy of the temperature calculations that the author has made for conventional tube structures, using the proposed solutions, was usually better than $\pm$ 10 per cent.

EXAMPLE OF HEAT TRANSFER CALCULATIONS — THE 6L6G PLATE AND SCREEN-GRID TEMPERATURES

The following examples illustrate possible solutions of the heat transfer equations for determining type 6L6G plate and screen-grid temperatures.

Operating Conditions: E_f = 6.3 volts, P_f = 5.67 watts

P_b = 18 watts

P_{c2} = 1.25 watts

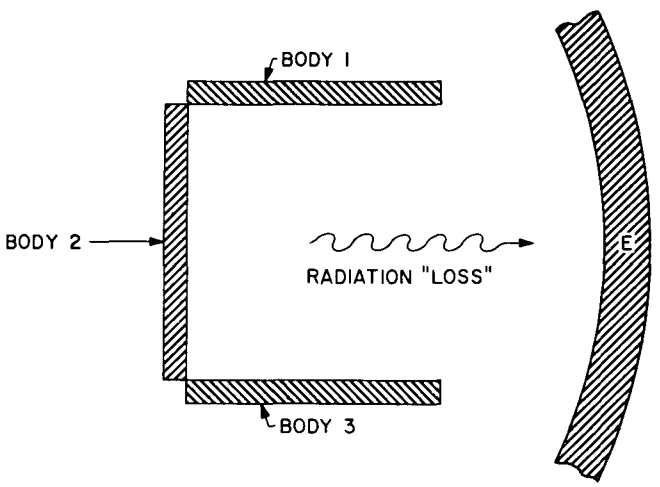


Figure 18. Cross Section of a "U" Surface

Solution for the Average Plate Temperature

The simplified heat transfer circuit for the determination of the average plate temperature is shown in Fig. 33. Because the plate temperature lies between that of the cathode and that of the enclosure, heat flow q_1 plus the 18 watts of plate dissipation must (by a nodal analysis of the circuit) equal the sum of heat flows q_2 and q_3. Because the heat flow through the stem lead q_3 is in the order of tenths of a watt (which is small compared to the 18-plus watts that the plate must lose), it will be neglected. The problem will be solved by assuming different plate temperatures and calculating the heat interchanges until the condition

is met. $$q_1 + 18 = q_2 \quad \text{(watts)}$$

The heat interchange equations are:

$$q_1 = \sigma A_k (T_k^4 - T_p^4) \frac{(1 - F_{k1} - F_{k2})}{(1/\epsilon_k) + (A_k/A_p)[(1/\epsilon_p) - 1]} \quad \text{(from 41)}$$

and

$$q_2 = \sigma A_p (T_p^4 - T_e^4) \epsilon_p \quad \text{(from 32)}$$

The values of the different parameters are:

T_k = 1060 K (true temperature)

T_e = 160 C = 430 K (from Fig. 35, which gives a hot-spot temperature of 185 C for a ST-16 bulb with pressed stem at 25 watts input; a slightly lower average temperature is assumed)

ϵ_k = 0.30

ϵ_p = 0.85

F_{k1} = 0.16 (from Fig. 14; 32 TPI of 3.3-mil wire)

F_{k2} = 0.13 (from Fig. 14; 32 TPI of 3.6-mil wire)

A_k = 1.7 cm^2

A_p = 21 cm^2 (assuming a smooth surface)

The solution for q_1 is

$$q_1 = 2.05\left[1.06^4\left(\frac{T_p}{1000}\right)^4\right]$$

If the average plate temperature is assumed to be 400 C (670 K), then q_1 = 2.17 watts, and q_2 = 2.17 + 18 = 20.2 watts. If this value is substituted into the equation for q_2, the resulting value of 695 K (425 C) is obtained. This value agrees reasonably well with the assumed 400 C. If a closer agreement is desired, a higher value than 400 C should be chosen and the calculations repeated. The final value for T_p will be close to 425 C.

Solution for the Average Screen-Grid Temperature

A simplified heat transfer circuit for determining average screen grid temperature is shown in Fig. 34. It will be assumed that the grid is hotter than the plate so that, by nodal analysis,

$$q_1 + 1.25 = q_2 + q_3 \quad \text{(watts)}$$

The heat-interchange equations are:

$$q_1 = \sigma A_k (T_k^4 - T_{g2}^4) \frac{F_{k2}}{(1/\epsilon_k) + (A_k/A_{g2})[(1/\epsilon_{g2}) - 1]} \quad \text{(from 34)}$$

$$q_2 = \sigma A_{g2} (T_{g2}^4 - T_p^4) \frac{F_{2p}}{(1/\epsilon_{g2}) + (A_{g2}/A_p)[(1/\epsilon_p) - 1]} \quad \text{(from 35)}$$

and

$$T_{g2} = T_L + q_3\left[(R_x/2) + (19/100)R_{FL} + (R_t/48N)\right] \quad \text{(from 38)}$$

where

$$T_L = T_s + (q_3/2)\,R_L$$

The values of the parameters are:

$A_k = 1.7\ cm^2$
$A_{g2} = 8.6\ cm^2$
$A_p = 21\ cm^2$
$T_k = 1060$ K
$T_p = 695$ K (from the preceding paragraph)
$T_s = 150$ C (estimated)
$F_{k2} = 0.13$
$F_{2p} = 0.16$
$\epsilon_k = 0.30$
$\epsilon_p = 0.85$
$\epsilon_{g2} = 0.85$ (carbonized)
$R_L = 1100$ C/watt (11 mm of 20-mil nipron)
$R_x = 216$ C/watt (9 mm of 40-mil nipron)
$R_{f4} = 875$ C/watt (36.5 mm of 40-mil nipron)
$R_t = 23,500$ C/watt (23.5 mm of 3.6-mil moly wire)
$N = 46$

Assume $T_{g2} = 800$ K; then $q_1 = 0.316$ watts. The sum of the screen-grid dissipation and q_1 is, then, 0.32 + 1.25 = 1.57 watts. The value of the grid-to-plate radiation of $q_2 = 1.11$ watts, and $q_3 = 0.45$ watts; their sum is 1.56 watts, which is close enough to the value obtained for q_1 to consider the 800 K (530 C) assumption valid.

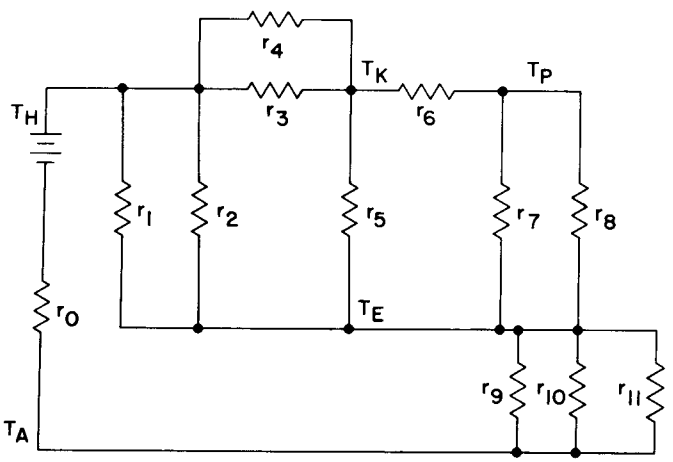


Figure 19. Heat Transfer Circuit of a Diode

If desired, the average siderod temperature may be calculated from the expression:

$$T_{sr2} = T_{g2} - q_3(R_t/48N)$$

The result for the 6L6G is 525 C.

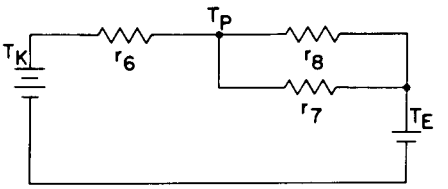


Figure 20. Simplified Heat Transfer Circuit of a Diode

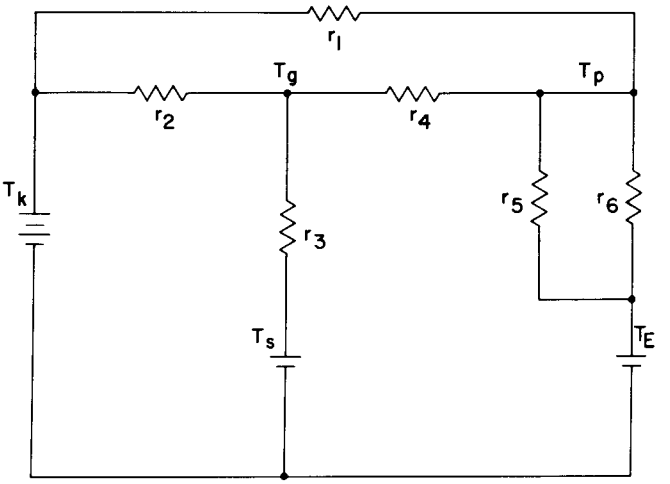


Figure 21. Heat Transfer Circuit of a Triode

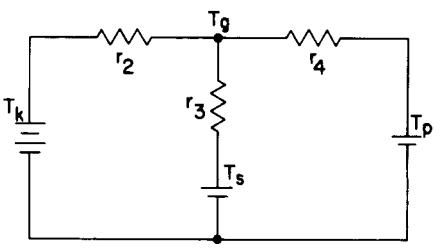


Figure 22. Simplified Heat Transfer Circuit of a Triode

LIST OF SYMBOLS

q	heat flow, heat interchange
q_b	black-body radiation
q_c	heat conduction in grid
E	total emissive power
E_b	total black-body emissive power
$E\lambda$	monochromatic emissive power
$E_{b\lambda}$	monochromatic black-body emissive power
λ	wave length
T	temperature
T_e	average enclosure temperature
T_g	average grid temperature
T_h	average heater temperature
T_k	average cathode temperature
T_L	siderod-to-stem-lead junction temperature
T_p	average plate temperature

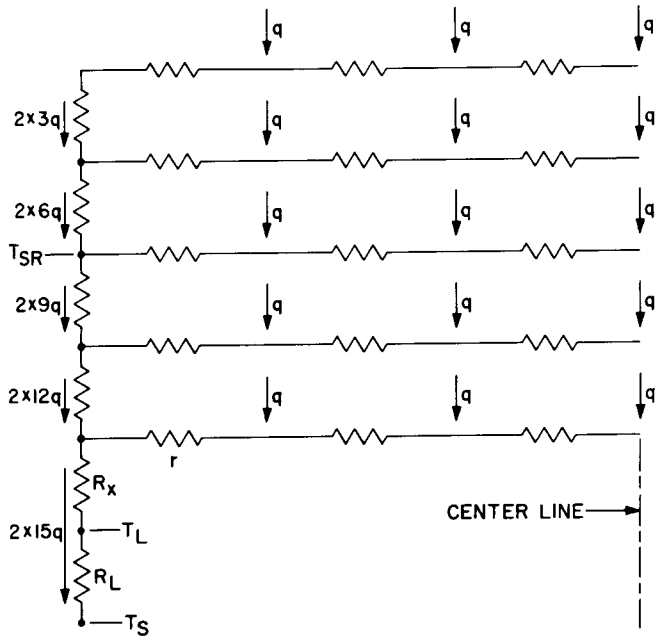


Figure 23. Thermal Resistances in One Side of a Grid Structure

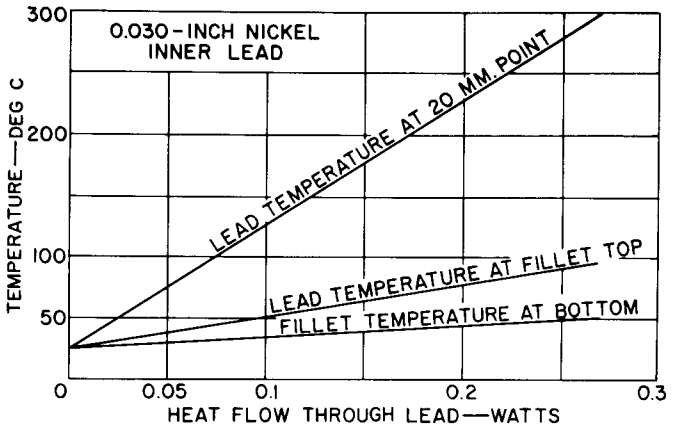

0.030-INCH NICKEL INNER LEAD
LEAD TEMPERATURE AT 20 MM. POINT
LEAD TEMPERATURE AT FILLET TOP
FILLET TEMPERATURE AT BOTTOM
TEMPERATURE—DEG C
HEAT FLOW THROUGH LEAD—WATTS

Figure 24. Stem Heat-Flow Characteristics for 1-1/4-Inch Button with Nickel Inner Leads

T_s	average stem (glass) temperature
T_{sr}	average grid-siderod temperature
ϵ_λ	spectral emissivity
ϵ_t	total emissivity
ϵ_e	total emissivity of enclosure
ϵ_g	total emissivity of grid
ϵ_k	total emissivity of cathode
ϵ_p	total emissivity of plate
K	thermal conductivity
K_m	mean thermal conductivity
R_r	thermal resistance
R_{fL}	thermal resistance of the "finished" grid length
R_L	thermal resistance of a stem lead

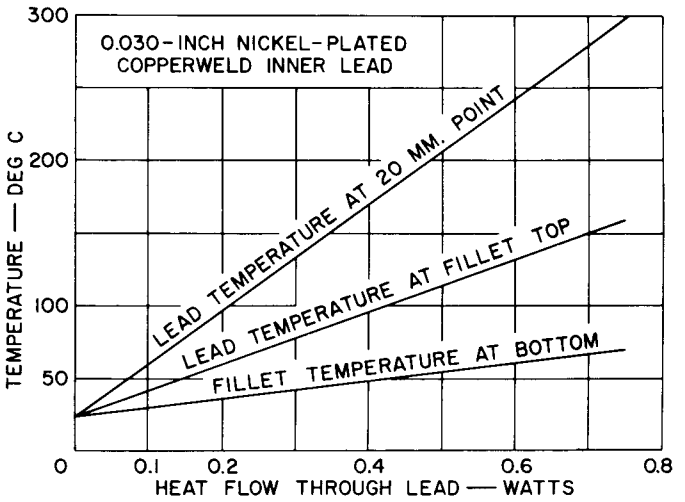

0.030-INCH NICKEL-PLATED COPPERWELD INNER LEAD
LEAD TEMPERATURE AT 20 MM. POINT
LEAD TEMPERATURE AT FILLET TOP
FILLET TEMPERATURE AT BOTTOM
TEMPERATURE — DEG C
HEAT FLOW THROUGH LEAD — WATTS

Figure 25. Stem Heat-Flow Characteristics for 1-1/4-Inch Button with Nickel Plated Copperweld Inner Leads

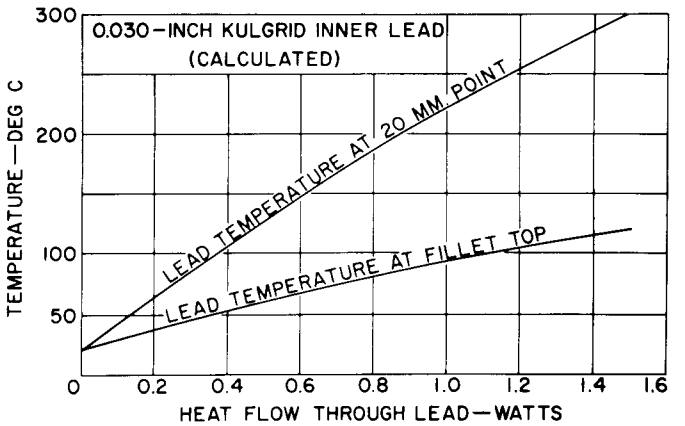

0.030-INCH KULGRID INNER LEAD (CALCULATED)
LEAD TEMPERATURE AT 20 MM. POINT
LEAD TEMPERATURE AT FILLET TOP
TEMPERATURE—DEG C
HEAT FLOW THROUGH LEAD—WATTS

Figure 26. Stem Heat-Flow Characteristics for 1-1/4-Inch Button with 0.030-Inch Inner Leads (calculated from the characteristic for 0.040-Inch Kulgrid inner-lead stem)

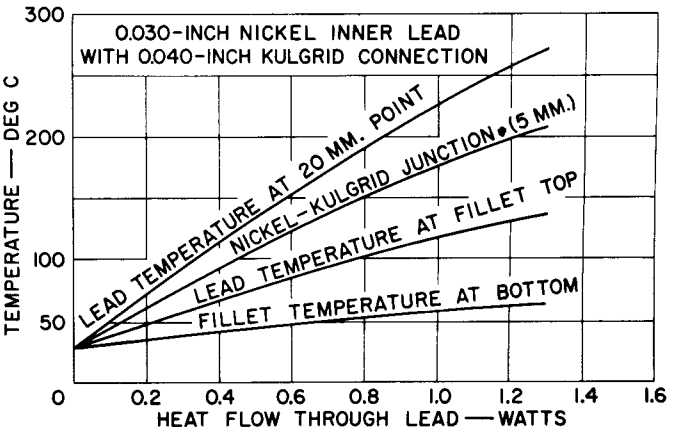

0.030-INCH NICKEL INNER LEAD WITH 0.040-INCH KULGRID CONNECTION
LEAD TEMPERATURE AT 20 MM. POINT
NICKEL-KULGRID JUNCTION (5 MM.)
LEAD TEMPERATURE AT FILLET TOP
FILLET TEMPERATURE AT BOTTOM
TEMPERATURE — DEG C
HEAT FLOW THROUGH LEAD — WATTS

Figure 27. Stem Heat-Flow Characteristics for 1-1/4-Inch Button with 0.030-Inch Nickel Inner Leads, 0.040-Inch Kulgrid Leads Welded to the Nickel at a Point 5 Millimeters Above the Fillet

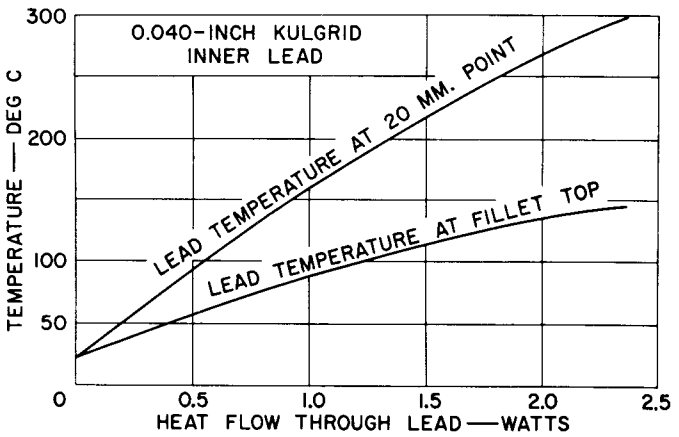


Figure 28. Stem Heat-Flow Characteristics for 1-1/4-Inch Button with 0.040-Inch Kulgrid Inner Leads

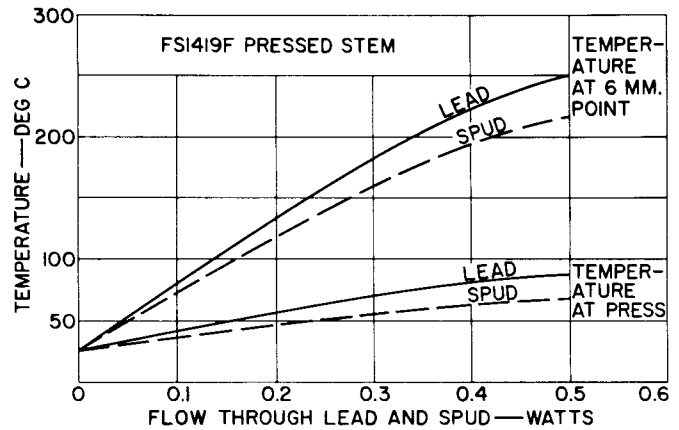


Figure 29. FS1419F Pressed Stem Heat-Flow Characteristics

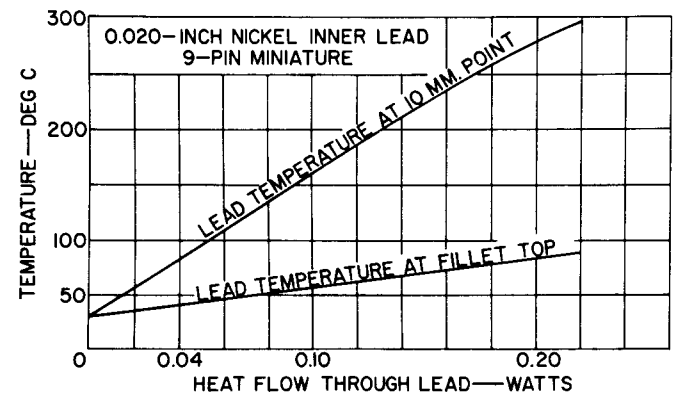


Figure 30. Stem Heat-Flow Characteristics for 9-Pin Miniature Button With 0.020-Inch Nickel Inner Lead

R_t	thermal resistance of one turn of lateral wire
R_x	thermal resistance of a grid leg
N	number of "finished" grid turns
A	area, cross-sectional area
A_m	mean cross-sectional area
A_e	enclosure surface area
A_g	"solid-surface" grid area
A_k	cathode-surface area
A_p	plate-surface area
F, $\bar{F}$, $\mathcal{F}$	view factors
x	length
D	diameter
σ	Stefan-Boltzmann constant
h	film coefficient of heat transfer
C_1, C_2	constants.

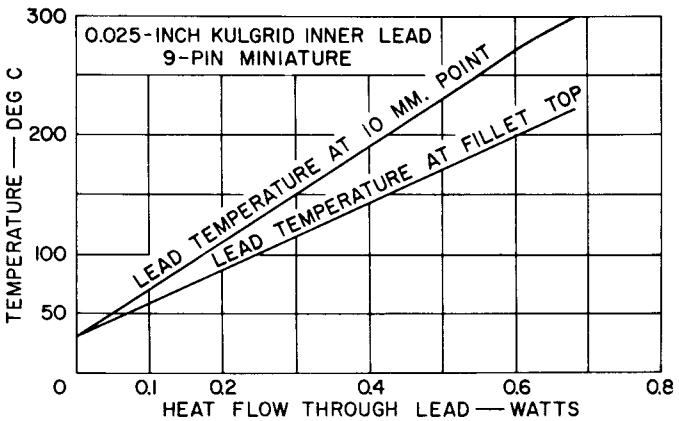


Figure 31. Stem Heat-Flow Characteristics for 9-Pin Miniature Button with 0.025-Inch Kulgrid Inner Lead

REFERENCES

HEAT TRANSFER BY STEADY UNIDIRECTIONAL CONDUCTION

1. Giedt, W. E., Principles of Engineering Heat Transfer, D. Van Nostrand, 1957
2. McAdams, W. H., Heat Transmission, 3rd Ed., McGraw-Hill, 1954
3. Obert, E. F., Elements of Thermodynamics and Heat Transfer, 1st Ed., McGraw-Hill, 1949

HEAT TRANSFER BY CONVECTION

See Refs. 1, 2, 3

THE PHENOMENON OF HEAT TRANSFER BY RADIATION THROUGH NONABSORBING MEDIA

See Refs. 1, 2, 3

4. Hottel, H. C., "Radiant Heat Transmission Between Surfaces Separated by Non-Absorbing Media," Trans. A.S.M.E., FSP-53-19b, p. 265, 1931
5. Temperature — Its Measurement and Control in Science and Industry, A.I.P. Symposium, Reinhold, 1941

SOLUTIONS OF HEAT TRANSFER EQUATIONS FOR RADIATION THROUGH NONABSORBING MEDIA

See Ref. 4

6. Hamilton, D. C., and W. R. Morgan, "Radiant-

Interchange Configuration Factors, "N.A.C.A. Tech. Note 2836

7. Hottel, H. C., "Radiant Heat Transmission," Mechanical Engineering, Vol. 52, No. 7, p. 699, 1930

THE APPLICATION OF THEORY TO HEAT TRANSFER PROBLEMS IN TUBE STRUCTURES

8. Code E1
9. Code E2
10. Code E3
11. Code E4

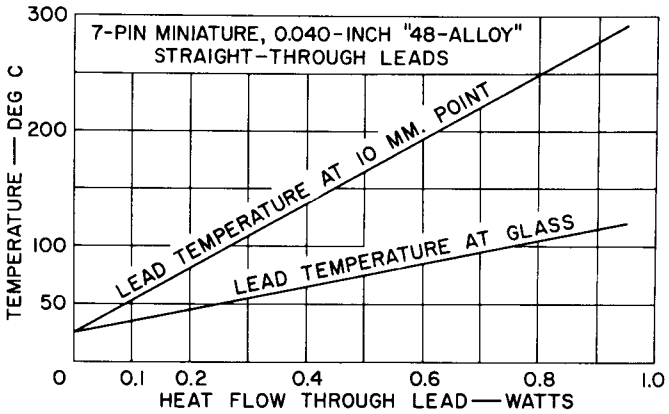


Figure 32. Stem Heat-Flow Characteristics for 7-Pin Miniature Button with 0.040-Inch "48-Alloy" Straight-Through Leads (Copper Plated), No Fillets

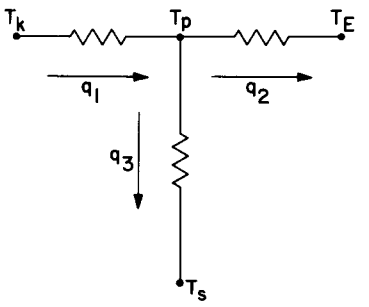


Figure 33. Simplified Heat Transfer Circuit for the Determination of the Average Plate Temperature

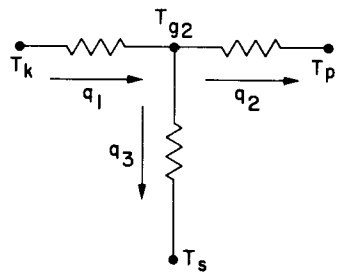


Figure 34. Simplified Heat Transfer Circuit for the Determination of the Average Screen-Grid Temperature

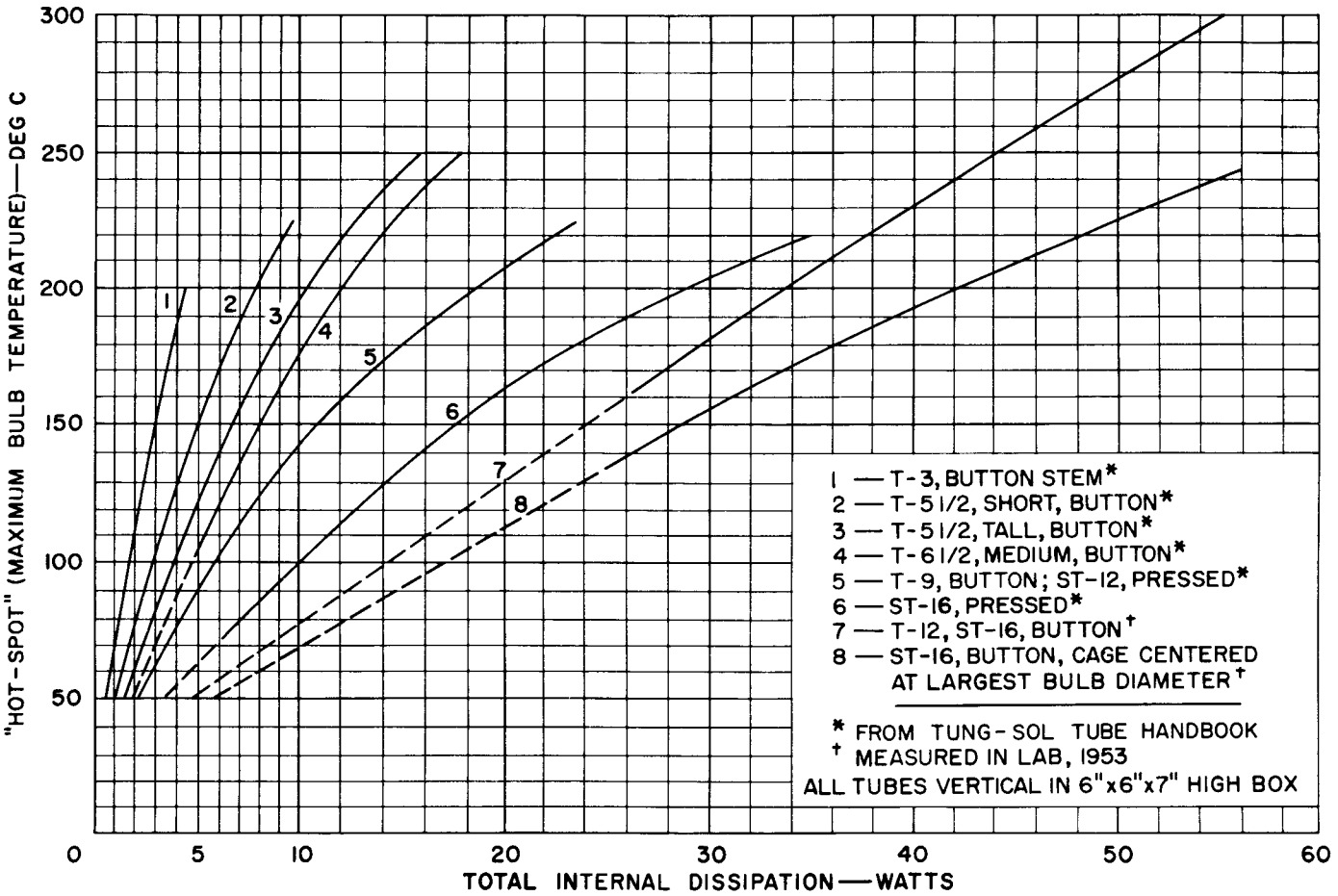


Figure 35. Maximum Bulb Temperature Versus Internal Dissipations for Various Bulbs

Contact Potential and Grid Currents

E. R. Schrader

Harrison

The importance of the control grid in a receiving tube stems from its ability to produce large changes in cathode current with comparatively small changes in potential. Anything that tends to shift the grid potential to a value other than that purposely applied can lead to poor performance of the tube in the circuit in which it is being used.

The purpose of this article is to discuss the origin of the more important factors which alter the effective control-grid potential (hereafter referred to simply as grid potential, or bias voltage) and to describe the means by which it is measured.

There are two ways in which the grid can influence the currents to the other electrodes other than that caused by the application of the desired signal potential. The first depends on the fact that the potential at the grid surface is invariably different than that applied to the grid wire metal. This potential difference must be taken into account when the effect which the grid wires exert upon the electric fields within the tube is considered. The second way in which the desired grid signal can be distorted is by the existence of unwanted grid currents that cause a voltage difference across any resistor in the grid circuit. Thus, one microampere of grid current will cause a one volt difference across a one megohm grid resistor.

This article is in two parts: The first part discusses electrode surface potential (work function) and contact difference of potential; the second part analyzes grid currents such as those caused by grid emission, gas, and leakage and the effect of these upon the grid current vs. grid voltage characteristics.

WORK FUNCTION AND CONTACT POTENTIAL

WORK FUNCTION

The movement of electrons from one electrode to another in a vacuum tube is determined by the resultant electric field produced by the applied electrode potentials, the electrode surface potentials, and any free charges which exist (space charge). The field at the surface of an electrode arises because an electron at a very short distance from the surface of the conductor in a vacuum is subjected to a combination of forces which tend to prevent its escape from the parent metal. [1, 2, 3] The electron may leave the surface of the metal but it will return unless it possesses enough energy to overcome these restraining forces. This behavior is comparable to that of a ball which is propelled up an incline and rolls back because it has not been given enough energy to make it go over the top. When a metal is heated, the internal electrons share the thermal energy in varying amounts, some getting enough energy to permit them to escape into the vacuum. The work function of a metal (analogous to the height of the incline) is the total energy which an electron at a particular reference energy level within the metal (called the Fermi level) must possess to escape completely from the metal. This energy is usually expressed in units of electron volts. A plot of the distance out from the surface to which electrons having a given amount of initial energy will go is called an energy profile diagram. Such a diagram is shown in Fig. 1. The horizontal axis represents the distance through the emitter and the vacuum; the vertical axis represents the energy of the electrons. Only those electrons receiving thermal energies greater than the work function can escape; all others return to the emitter. The energy profile diagram of a semiconductor surface is more complex.

CONTACT POTENTIAL

An electron in the space between two electrodes is affected by the surface potentials of both electrodes if they are so close that their energy profiles overlap. In most cases, the electrodes of receiving tubes are much too far away from each other for such overlapping to be the case. The actual field seen by an electron between two relatively distant electrodes can be indicated diagramatically by drawing the work functions as inclines with a straight line joining their upper rims, the relative position of each work function being additionally positioned by any externally applied fields. A space charge due to a high density of electrons will cause an alteration of the existing potential distribution. When the two electrodes are connected in a circuit in series with a battery, the whole energy level of each electrode is shifted relative to the other by the potential difference of the battery. Because the battery is connected by leads directly to the metal of the electrodes (bypassing the surface, or work-function, effects), it is the Fermi levels of the electrodes and not the upper rims of the work-functions which differ by the battery potential.

Fig. 2 shows how the potentials of two separate electrodes vary with respect to each other as they approach

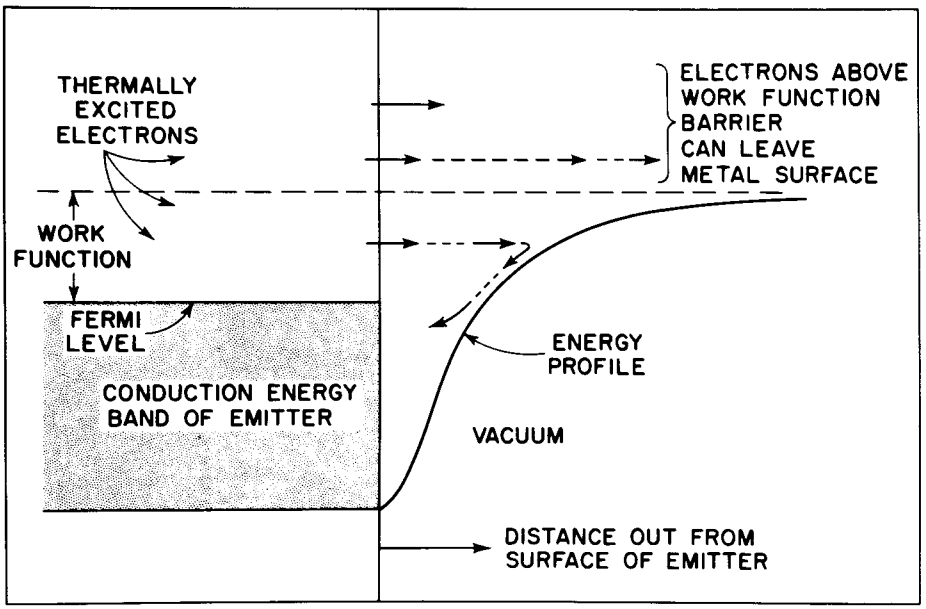


Figure 1. Simplified Energy-Profile Diagram of a Metallic Thermionic Emitter in a Vacuum in the Absence of an Applied Field

closely and come into intimate contact.[4] Fig. 2A shows the energy profiles of two metals approaching to make contact at their near surfaces. If no external sources of potential are assumed, the electric field in the gaps is zero, causing the Fermi levels in the metals to differ in energy level by the difference in work functions. Fig. 2B shows the energy profiles of the same two metals in contact. Upon close approach and contact at the near end, electrons flow from metal A to metal B until the internal Fermi levels are equal. This serves to separate the work-function tops in the far gap and causes a contact difference of potential and a resultant electric field gradient. Any electrons which may be in this gap (i.e. due to thermionic emission from either metal) will react to this field just as if it existed by virtue of an externally applied potential difference (battery) between the separated electrodes. Fig. 2C shows how the energy levels adjust for the use of two electrodes and an intermediate shorting wire when all are at room temperature. The same metals as in Figs. 2A and 2B are in mutual contact with a third metal which has a different work function. The work function of metal C has no effect; this metal acts only as a carrier of electrons when the Fermi levels are all made equal in value. The resulting contact potential between A and B is the same as in Fig. 2B. Metal C, then, is similar to the case of a shorting wire between two tube electrodes.

Note that the electrons in the electrodes and wire are at the same energy level. This situation, then, is typical of that existing when two electrodes of a tube are at zero potential relative to one another. However, it is to be noted that the surface energies of the metals differ by the difference in their work functions where they are not in contact. In other words, because the surfaces are in contact at one place (directly or through at intermediate conductor), the other surfaces at the gap are at a potential difference which is equal to the difference in their work functions. This statement is the reason for assigning the term "contact difference of potential" (usually shortened to "contact potential") to the work-function difference.

What has been said so far applies to pure metal emitters but not necessarily to the oxide-coated cathode which is the emitting electrode most often used in receiving tubes. The mechanism of emission from the oxide-cathode is more complex than that from a simple metal and is fully discussed in the articles by Meltzer and Widell. The work function of an oxide cathode is a function of its operating temperature and the current emitted.[5,6] For most such measurements on commercial electron tubes, however, the influence of current and temperature is considered to be a second-order effect. In this discussion of surface potentials, it is assumed that the simplified picture of work function applies. It is also assumed that the contact potential of interest is that between the cathode and the adjacent electrode (usually the control grid).

THREE METHODS OF MEASURING CONTACT POTENTIALS

The methods used to observe the electric properties of a tube to indicate electrode surface effects are in three general categories. In the first method, a potential is determined graphically by the extrapolation of two well-divided regions of the grid-current-vs.-grid-voltage characteristic curve taken at low cathode temperatures. This procedure yields what will be considered a true contact-potential measurement. In the second method, a potential is determined from the applied voltage required to produce a chosen value of collector current. In the third method, a potential is determined from the reading of a high-impedance voltmeter connected between grid and cathode.

The oxide cathode, being a semiconductor, exhibits a thermovoltaic effect which modifies the work function and must be corrected for in more exacting work. The

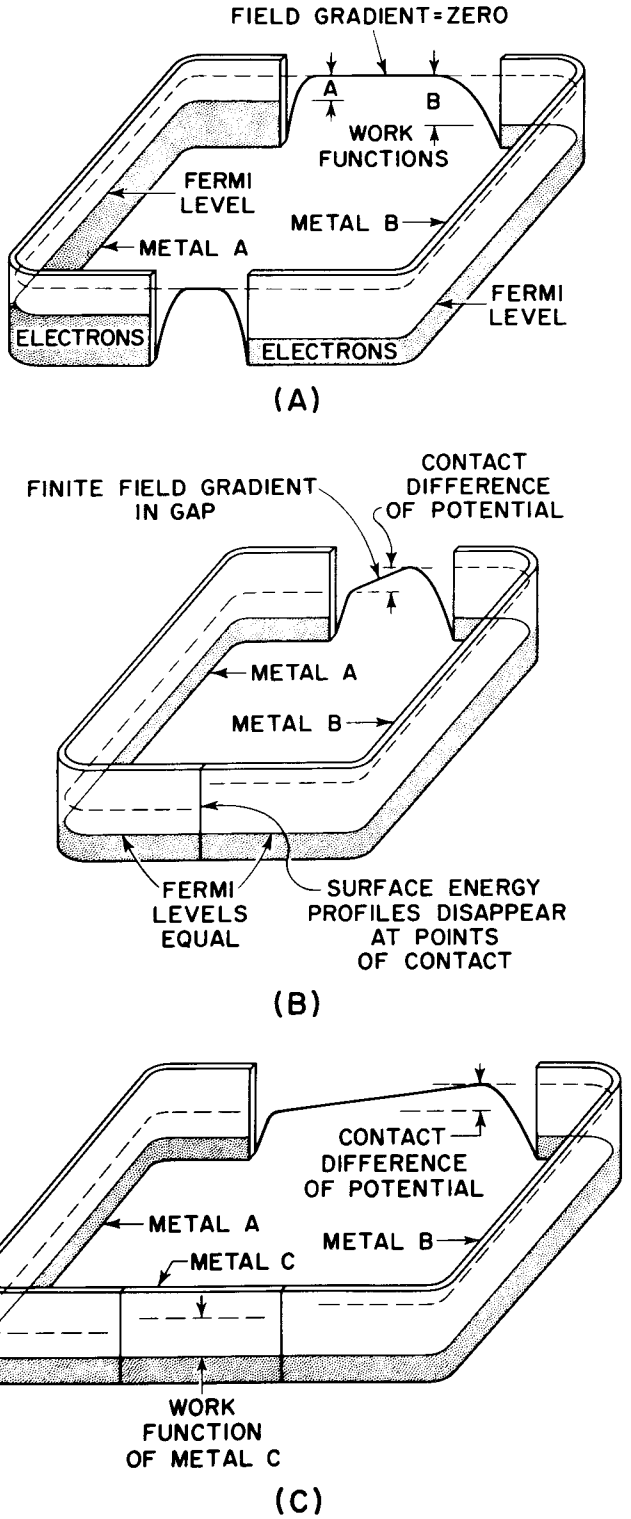


Figure 2. Illustration of the Concept of Contact Potential Between Two Metals: (A) case of two approaching metals separated by two different gaps; (B) case of two approaching metals where one gap has closed but other gap remains; (C) case of two approaching metals where one gap has been closed by a third metal, but other gap remains

magnitude of this correction is of the order of 0.003 volt per degree of temperature difference across the coating.[7] There is also a small effective voltage due to the initial velocity of the electrons.[8] The three methods and their significance are described in greater detail below.

The Intersection Method for Measuring Contact Difference of Potential (Method I)

In the intersection method, the cathode is heated to a temperature of approximately 600 to 700 K by operation of the heater at about one quarter of its rated voltage (e.g., 1.5 volts for a 6.3 volt heater). To minimize the effect of the plate (assuming the tube is a triode), the plate is connected to the cathode. The logarithm of the grid current is then plotted as a function of the applied grid voltage. The plotted grid-characteristic curve usually contains three clearly defined regions that represent the effects of the retarding field, the space-charge field, and the saturating field.

The technique by which such a curve may be used to determine true contact potential is illustrated in Fig. 3. Assume that the cathode has a work function of 1 electron-volt and the grid has a work function of 2.5 electron volts. (It is generally true that the grid has a higher work function than the cathode). In Fig. 3, point (A) gives the energy-vs.-distance configuration corresponding to an applied grid voltage of -0.25 volt which causes an assumed grid current of 0.001 microampere. Even though the grid current is quite small, the electrons form a space-charge field, with the result that the potential distribution between the grid and the cathode is not linear. The space-charge field, however, does not form a barrier higher than that of the grid work function, so that the latter plus the applied field are the only factors limiting the current. Only electrons that have thermal energies greater than the work function of the cathode plus the applied -0.25 volt can reach the grid to provide the 0.001 microampere of grid current.

Point (B) shows the profile for zero applied voltage, i.e., the Fermi levels of the cathode and the grid have the same energy value. Under these conditions, the only retarding field affecting the electrons is that produced by the difference in work functions, i.e., by the true contact potential. Some of the lower-energy electrons now can reach the grid with the result that the grid current increases. When the grid is made positive with respect to the cathode, as at point (C), the energy profile approaches the horizontal, and at a certain point the space-charge field begins to become important in limiting the current. Up to this point the grid-characteristic plot is a straight line because in the retarding-potential region the collector current for an oxide cathode (like that from a pure metal emitter) is generally a logarithmic function of applied voltage, as predicted by theory.[9] The interposition of the space-charge barrier at this point, however, tends to reduce the current, with the result that the curve starts to round off, as at (D).

Now if the grid is made sufficiently positive, as at (E), both the grid work function and the space-charge barrier no longer have a limiting effect on the current

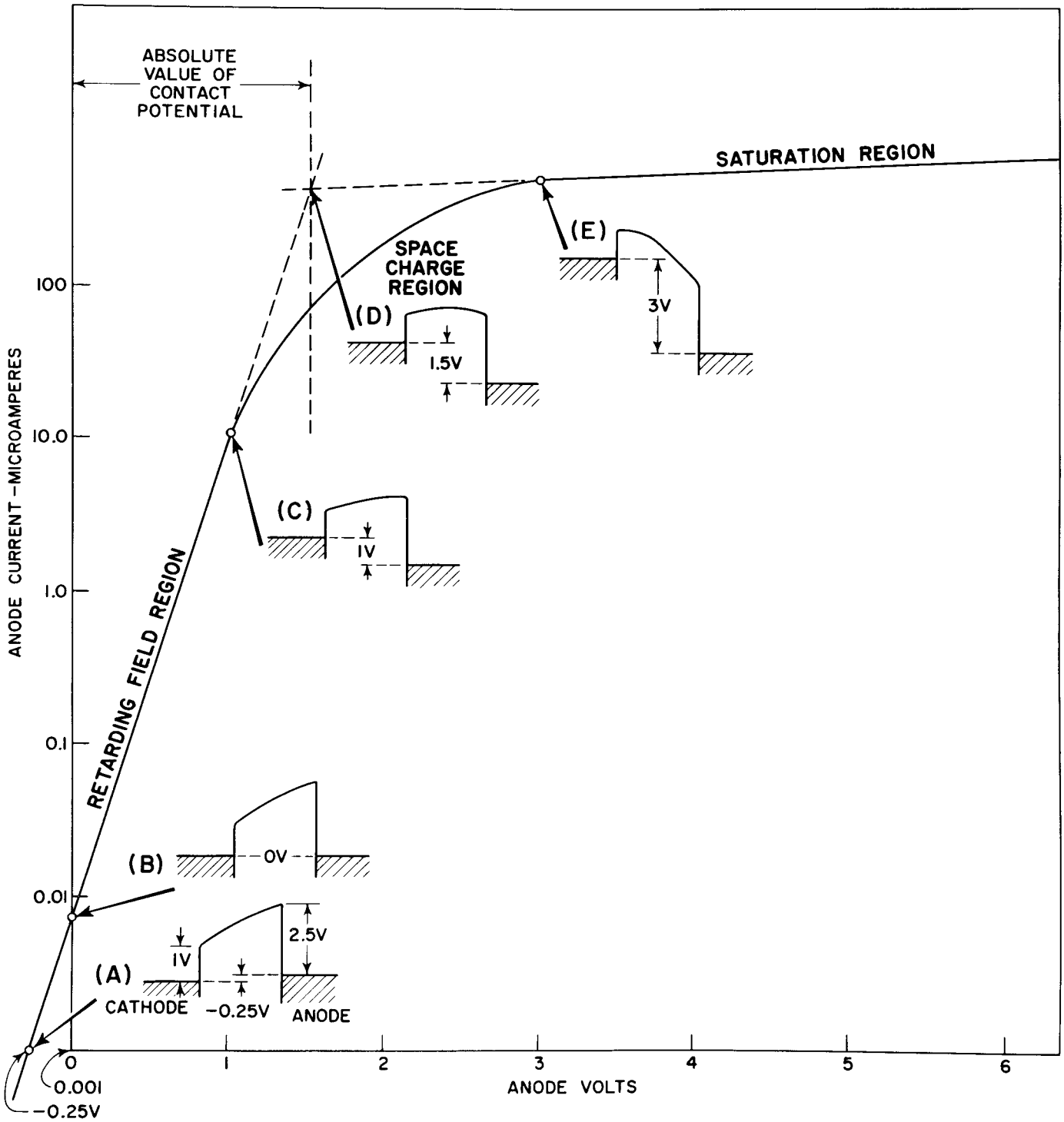


Figure 3. Idealized Diagram Showing the Principal Regions of a Grid-Characteristic Curve and the Determination of Contact Potential by the Intersection Method

and, as a result, all the electrons emitted are then collected. However, the current in this saturation or temperature-limited region of the grid-characteristic curve does not remain constant but increases slightly with applied voltage due to a slight change in the cathode work function by the applied field.

If no space-charge barriers were present and if the grid and cathode work functions were uniform, the retarding-field region of the curve would change abruptly into the saturation region at the applied grid voltage which sets the upper rims of the cathode and grid work function at the same level at (D). This applied voltage is numerically equal to the difference in work functions, and therefore, is the true contact potential—the desired measurement value. Actually, because there is always some rounding of the curve, this voltage is given by the intersection of the extrapolation of the retarding and saturated regions of the curve. When circuits are being considered, the contact potential can be con-

sidered negative because it acts in the same manner as a negative grid-bias voltage would for the usual case where the grid work function is larger than the cathode work function.

As mentioned previously, a determination of contact potential is made with the heater voltage at approximately one-quarter of the rated value. The reason for this choice can be seen from Fig. 4 where a typical family of contact potential curves are drawn with cathode temperature as a parameter.[10] As the cathode temperature is increased, the number of electrons which pass the work-function barrier increases. The current I emitted past the work-function barrier is given approximately by the expression[11]

$$I = Ae^{(e/k)(\phi_k/T)} \qquad (1)$$

where:

A = a constant which can be empirical or, in some cases, can be theoretically derived and contains a dependence upon T
e/k = the charge of an electron divided by the Boltzmann constant to give 11,606
ϕ_k = the cathode work function in electron volts
T = the cathode temperature in degrees Kelvin

Each rise in temperature, then, is accompanied by an increase in current at each successive voltage increment as applied along the contact potential curve, effectively raising the complete curve. To a first approximation, however, the intersection point which determines the contact potential remains the same because the dependence of contact potential on temperature is relatively minor. Therefore, the actual temperature chosen for this measurement is not critical. However, it is evident from Fig. 4 that the extrapolation of the retarding field and temperature limited lines is made much more difficult at the higher temperatures because of the extreme rounding of the curve due to the added space charge. In addition, because higher cathode temperatures cause an increase in the temperature of all other tube parts, it is more likely that extraneous effects, such as mica or stem leakage, will distort the current readings. Because the emission current changes exponentially with temperature, the use of high temperatures will result in an increase in collector (usually the grid) dissipation with additional harmful effects. Experience has shown that a cathode temperature which results from applying approximately one-quarter of the heater voltage is a safe value.

A convenient method for obtaining a meter indication of the contact potential has recently been divulged by A. Dingwall.[12] His method makes use of the fact that the point of intersection of the retarding-field boundary and the saturated-region boundary of the curve occurs at approximately the same voltage as does the fastest change of the slope of the rounded portion (the maximum value of the second derivative). By sweeping the collector with a sawtooth voltage relative to the cathode, converting the resulting current to its logarithmic value, and electronically taking successive derivatives, an output is derived which is proportional to the voltage at which the maximum second derivative occurs. With this method, contact potential can be determined in the time it takes the tube to reach its equilibrium temperature.

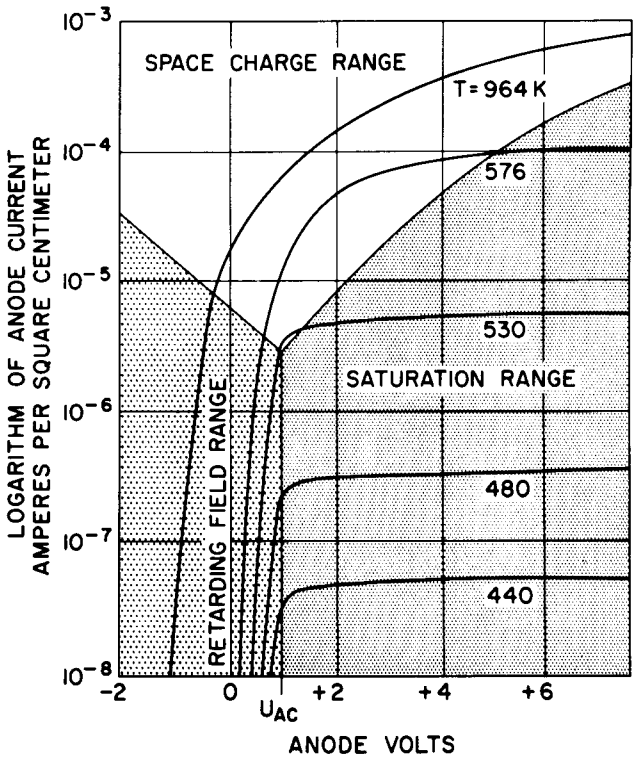


Figure 4. A Typical Family of Contact-Potential Curves with Cathode Temperature as a Parameter. (Based on Ref. 1, Vol. II, p. 28.)

Significance of the Contact Potential Curve. There is a danger in receiving-tube engineering practices of using the concept of contact potential erroneously. The danger stems mainly from the fact that the contact potential gives only a difference in the cathode and collector electrode work functions and not the absolute value of either. A change in contact potential from one condition to another in a receiving tube gives no information as to which of the work functions was responsible for the change. If the control-grid work function changes, there is a different effect on the space-charge-limited operation of a tube than if the cathode work function changes by an equal amount in an opposite direction. Either change results in the same change in contact potential. It is not true (as is generally supposed) that a change in contact potential is always equivalent to an equal change in grid bias voltage. The latter statement is true only when the change in contact potential is due to a shift in the grid work function alone. Where formulas appear in which the contact potential is legitimately inserted in such a way as to appear as a bias voltage (such as some versions of the Child-Langmuir equation), there are other terms in the equation that depend upon cathode emission (and, thus, the cathode work function) which tend to compensate for that part of the contact potential shift due to the cathode-work-function change.

If, in a tube problem, care is taken to get the data necessary to plot a contact-potential curve, more in-

formation is available from these date than just the value of contact potential alone. Referring back to Fig. 3, the current collected on the vertical portion of the curve (retarding-field portion) is limited by the collector work function as well as by the potential applied to the collector. A decrease in the collector work function would allow more current at all points on the retarding field curve, effectively raising it, and giving it the appearance of moving to the left. At any given arbitrary reference current on this curve, the voltage change as the curve shifts is directly equal to the change in the collector work function which caused the shift. Therefore, for an individual tube, any changes of grid work function can be observed by noting the magnitude and direction of the shift of the retarding-field curve. This condition is illustrated in Fig. 5A. In Fig. 4, similar shifts resulted if the cathode temperature at which the measurement was made was changed. Therefore, even though the measured value of contact potential itself is not a critical function of the cathode temperature, the information abstracted from the curve concerning the grid work function can be in error if all measurements are not taken at the same cathode temperature.

In an analogous way, changes in the horizontal (temperature-limited) portion of the contact potential curve

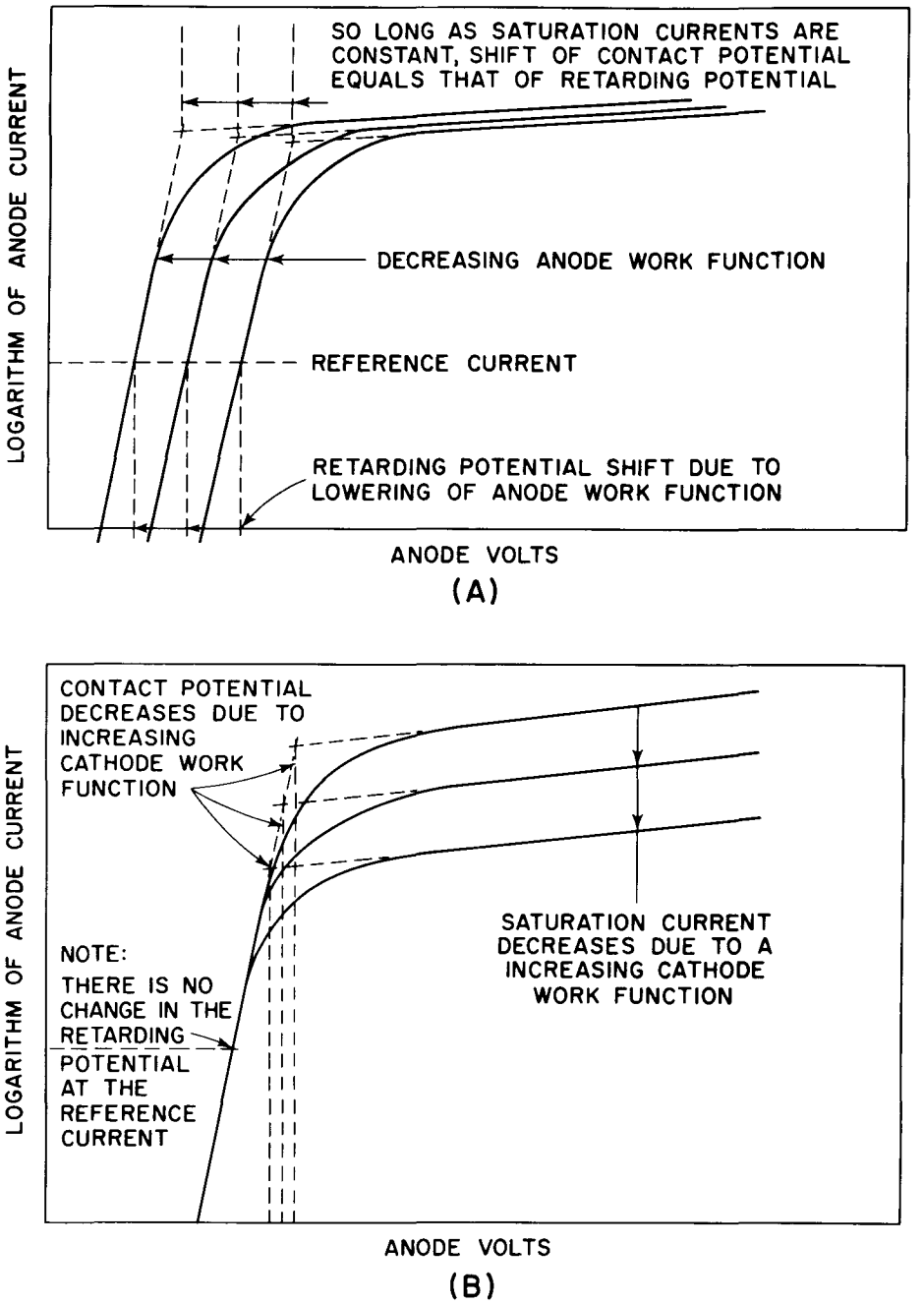

SO LONG AS SATURATION CURRENTS ARE CONSTANT, SHIFT OF CONTACT POTENTIAL EQUALS THAT OF RETARDING POTENTIAL
DECREASING ANODE WORK FUNCTION
REFERENCE CURRENT
RETARDING POTENTIAL SHIFT DUE TO LOWERING OF ANODE WORK FUNCTION
LOGARITHM OF ANODE CURRENT
ANODE VOLTS
(A)
CONTACT POTENTIAL DECREASES DUE TO INCREASING CATHODE WORK FUNCTION
SATURATION CURRENT DECREASES DUE TO A INCREASING CATHODE WORK FUNCTION
NOTE:
THERE IS NO CHANGE IN THE RETARDING POTENTIAL AT THE REFERENCE CURRENT
LOGARITHM OF ANODE CURRENT
ANODE VOLTS
(B)

Figure 5. Effects of Work-Function Shifts: (A) illustrating the shift in the retarding potential at some reference current due to grid-work-function shifts; (B) illustrating the shift in the saturation currents due to cathode-work-function shifts

are determined only by changes in the cathode work function, where an upward shift of the curve results from a decrease in the cathode work function as shown in Fig. 5B. At some reference grid voltage then, the change in emission current which results from a change in cathode work function can be determined. Because the shift is along the current axis, it does not give a number which directly determines the work-function shift. The latter must be calculated from an expression, such as Eq. (1), where a value for the coefficient A must be derived from other measurements, and the equation solved for ϕ_K. However, this procedure is not generally necessary since it is usual to judge the emission capability of the cathode by other methods (i.e., pulse techniques). The chief value of observing this portion of the contact-potential curve is to determine the direction of any change of cathode work function — the direction of change being a valuable aid in determining whether tube-processing steps have adversely affected the cathode. Once again, it should be mentioned that these methods give true results only while the temperature of the emitter is held constant.

The Method of Measuring a Potential to Produce an Arbitrary Reference Current (Method II)

It was mentioned previously that the horizontal shift of the retarding-field portion of the contact-potential curve is equal to the change in collector work function provided the cathode temperature is held constant. An additional necessary assumption is that the resistance of the cathode coating does not change and thereby cause an unsuspected voltage change in the system. If, under these ideal conditions, changes in collector work function are to be monitored, it is necessary only to observe changes in the voltage needed to give some arbitrary reference current on the retarding-field line. For tubes containing cathodes and interelectrode spacings such as are usual in miniature or larger-size receiving tubes, a reference current of 0.1 microampere is usually in the retarding-field region. However, a higher value, such as 1.0 microampere, is sometimes more convenient because less care is necessary to avoid the pick up of extraneous signals.

The same principle can be used at higher cathode temperatures, the result being that the greater number of available high-energy electrons will require a larger retarding field (a more negative collector voltage) to be applied in order to keep the actual collector current at the reference value. Thus, grid work-function shifts can be monitored using rated as well as lower heater voltages. In such cases, the polarity of the collector voltage is almost always negative. The advantage of using rated heater voltages is that the determination of the retarding bias voltage needed to give the reference current can be combined with other conventional tests without special procedures and preheaters. However, as will be more fully explained later, there can be a contribution from gird-cathode leakage and grid-emission currents at the higher heater temperatures. These currents are due to electrons which leave the grid and flow in a direction opposite to those arriving from the cathode. As a result, a different potential is needed to give the reference current and, therefore, changes in the resulting retarding potential are no longer strictly dependent on only changes in work function.

The voltage necessary to give some chosen reference current is often erroneously used as, and named, contact potential. Referring to Fig. 5A, it can be seen that any change in this voltage is indeed the same as the contact-potential change, but only as long as the cathode work function remains constant. For the case illustrated in Fig. 5B, it is obvious that the voltage to give the reference current does not change. In any case, the actual value of the collector voltage necessary to give a reference current is never the same value as the voltage at the intersection point, the latter being defined as the true contact potential.

Also, because the choice of a reference current is, within limits, an arbitrary one, the absolute value of the measured voltage has meaning only in terms of its changes. Retarding potential[13] is another, and more correct, term used for this measurement in the tube industry. To avoid the restrictive use of so general a term, it should be accompanied by the reference current, as for example, the "0.1-microampere retarding potential," or "the retarding potential for 0.1 microampere." In addition, it should clearly be understood what heater voltage is being used for the measurement. Accordingly, the symbol E_{RH} is used to specify the voltage necessary to get a reference grid current at rated heater voltages (the subscript R is used for retarding, and H for a hot cathode). The symbol E_{RC} is used for a heater voltage of approximately one-quarter of the rated value (i.e., 1.5 volts for a 6.3 volt tube, the subscript C indicates a cool cathode). Because plate current usually produces gas grid currents which materially alter the grid-current-vs.-grid-voltage characteristic curve, the retarding-potential readings are usually determined with the plate floating or grounded if the valves obtained are to be used as indications of grid-work-function changes.

To summarize, Fig. 4 shows that E_{RH} will generally be a negative voltage, E_{RC} a positive voltage, and E_{CP} also positive but always higher than E_{RC}. If only the grid work function is changing, each of these values will shift by the same amount and in the same direction. However, E_{RH}, being a negative voltage, will show a change in absolute value opposite to the other two quantities. This latter fact is responsible for a great deal of confusion when two quantities which change oppositely in absolute value are used interchangeably. Thus, as an example, when it is stated that a given retarding potential is decreasing with tube life, it is very important that the measurement be fully defined.

The Method of Measuring Potential with a Grid Voltmeter (Method III)

If a dc voltmeter having a high resistance (about 10 megohms) is connected between the grid and cathode of a tube, and no other connection is made to the grid, a voltage difference will develop across the meter due to the grid current of the tube. Graphically, this voltage is determined by the intersection of the meter-resistance load line with the grid-current-vs.-grid-voltage curve. In method II, it was pointed out that the grid curve will shift with changes in grid work function. This

shift, of course, will also alter the point of intersection of the grid curve with the meter load line and, therefore, produce a change in meter reading. Thus, the meter reading can be used as a measure of grid-work-function shifts. However, any readings and changes in the readings depend upon the slope of the load line (i.e., the load line due to meter resistance) and are, therefore, only proportional to the magnitude of the shift. A graphical representation of this measurement is given in Fig. 6. The term commonly applied to describe it in the industry is developed bias (E_D). However, to have more meaning, the resistance which determines the load line should be included, such as "10-megohm developed bias," or better, "developed bias for 10-megohm load." Sometimes, the voltmeter is paralleled by an additional resistance to provide a different load line better suited to the particular test setup or a particular circuit condition.

As with other methods, the measurement should be made with no plate current in order to avoid the distortions due to grid gas current if the primary objective is to follow grid-work-function changes. However, as will be discussed later, measurements under the working conditions (i.e., with plate voltage applied) can yield developed bias-voltage information of a different significance.

GRID CURRENTS

The first part of this article described the origin of those electric fields within a tube due to the grid and cathode surfaces, and the means by which the effects of these fields are measured and categorized. This part of the article discusses the currents within a tube which are measurable at the grid — such currents are partially dependent upon the electrode surface potentials discussed earlier.

There are four main sources of grid current: (1) electrons which arrive at the grid by direct emission from the cathode, (2) electrons which leave the grid by direct primary emission from the grid, (3) electrons which leave the grid because of leakage paths on the insulators (micas and bulb stem) which support it, and (4) positively-charged gas ions which are formed by plate-current bombardment of the gaseous environment in the tube. Each of these grid currents has a unique grid-voltage dependence. Any method of analyzing for the magnitude of each effect depends upon separating and recognizing their characteristics.

CATHODE-TO-GRID EMISSION

Cathode-to-grid emission was partially discussed in the first part of this article; all of the grid-current-vs.-grid-voltage curves shown were assumed to be composed of electrons emitted from the cathode and collected on the grid, just as if the grid were the anode of a diode. The thermal energy imparted to the cathode electrons enables them to overcome slight retarding fields due to the grid potential and thereby give a grid current even when the grid is biased negatively. The

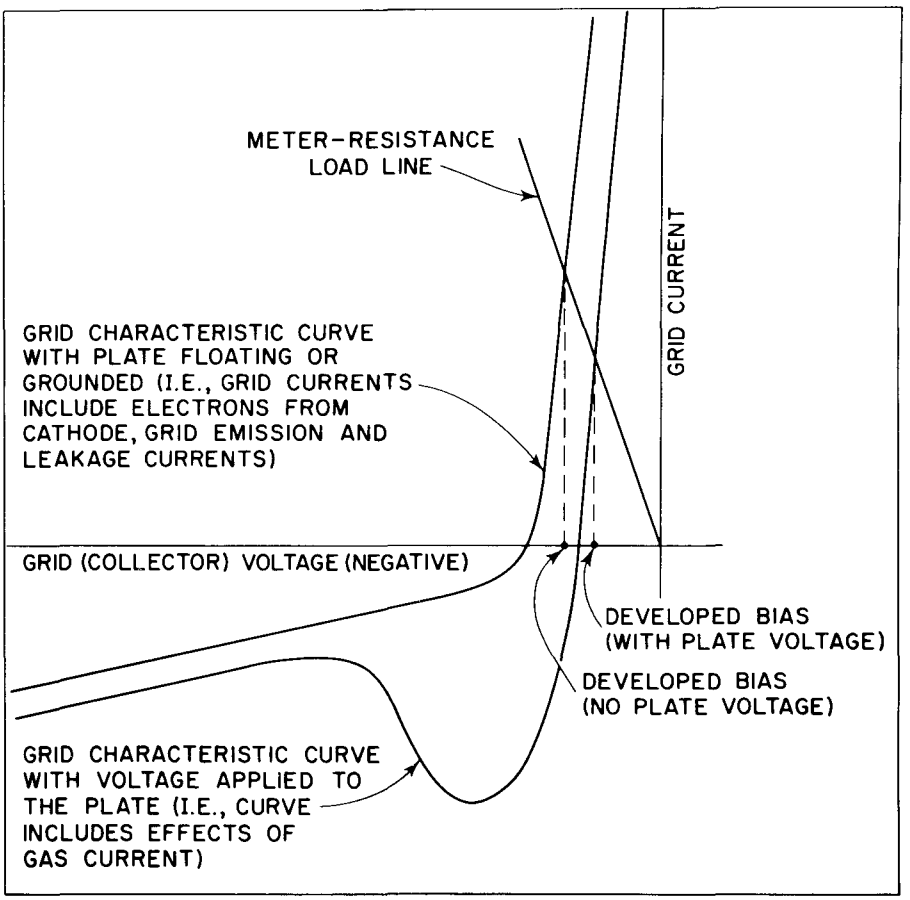

METER-RESISTANCE LOAD LINE
GRID CURRENT
GRID CHARACTERISTIC CURVE WITH PLATE FLOATING OR GROUNDED (I.E., GRID CURRENTS INCLUDE ELECTRONS FROM CATHODE, GRID EMISSION AND LEAKAGE CURRENTS)
GRID (COLLECTOR) VOLTAGE (NEGATIVE)
DEVELOPED BIAS (WITH PLATE VOLTAGE)
DEVELOPED BIAS (NO PLATE VOLTAGE)
GRID CHARACTERISTIC CURVE WITH VOLTAGE APPLIED TO THE PLATE (I.E., CURVE INCLUDES EFFECTS OF GAS CURRENT)

Figure 6. Illustrating the Source of Developed Bias Voltage Readings

characteristic curve, illustrated in Fig. 7A, shows that as the grid voltage becomes progressively less negative, the number of electrons entering the grid increases very rapidly.

PRIMARY GRID EMISSION

Like the cathode, the grid can emit electrons when it is hot. For a bare metal wire, the grid work function would be quite high (4 to 6 electron volts), and the emission current would be comparatively low at normal grid temperatures (600 to 800 K). However, many monolayers of sublimed cathode coating material contaminate the grid surface after tube processing and the work function is usually in the range of 1.5 to 2.5 electron volts. The characteristic grid-current-vs.-grid-voltage curve for grid emission in Fig. 7B is similar to the emission curve of Fig. 7A, except for the reversed polarity and the lower value of saturation current due to the lower grid temperatures and higher work function. At some small negative grid voltage, the cathode (collecting the electrons in this case) is relatively positive enough to draw over all of the electrons emitted from the grid; as a result, grid-emission saturation is quickly reached and the curve levels off. The value of this saturation current is the grid emission, generally about 0.001 to 0.1 microampere, but sometimes as high as one microampere. Actually, there is a slight slope to the saturation line but this effect is usually insignificant compared to that due to the leakage characteristics to be described later. Because grid emission is affected by the amount of sublimed cathode material, the degree of activation or poisoning, the grid wire material itself, and the grid geometry, it is a difficult quantity to control or predict. Further aspects of this problem will be discussed later.

LEAKAGE CURRENTS

The mica parts used to support tube electrodes and the glass stem used to support and provide electrical contacts for the cage are, of themselves, usually excellent insulators. However, during tube processing and activation, metallic and semiconducting films form and provide a leakage path between lead wires and between electrodes. When operating voltages are applied to the electrodes small currents will flow. If these currents are in the grid circuit, they can be harmful to the tube operation.

Because the grid is generally at the lowest tube potential, electrons flow from the grid to the cathode and to the plate. This flow results in characteristics such as are illustrated in Figs. 7C and 7D. The grid-to-cathode electron flow is zero when the grid potential is zero, and the grid-to-plate current would be zero only if the grid were at the plate potential. As is usually the case, the leakage resistance appears ohmic and there is a straight-line relationship between voltage and current.

GAS ION CURRENTS

Due to gases which exist within the tube envelope after processing the tube, and other gases which come from tube parts during tube operation, there will always be a

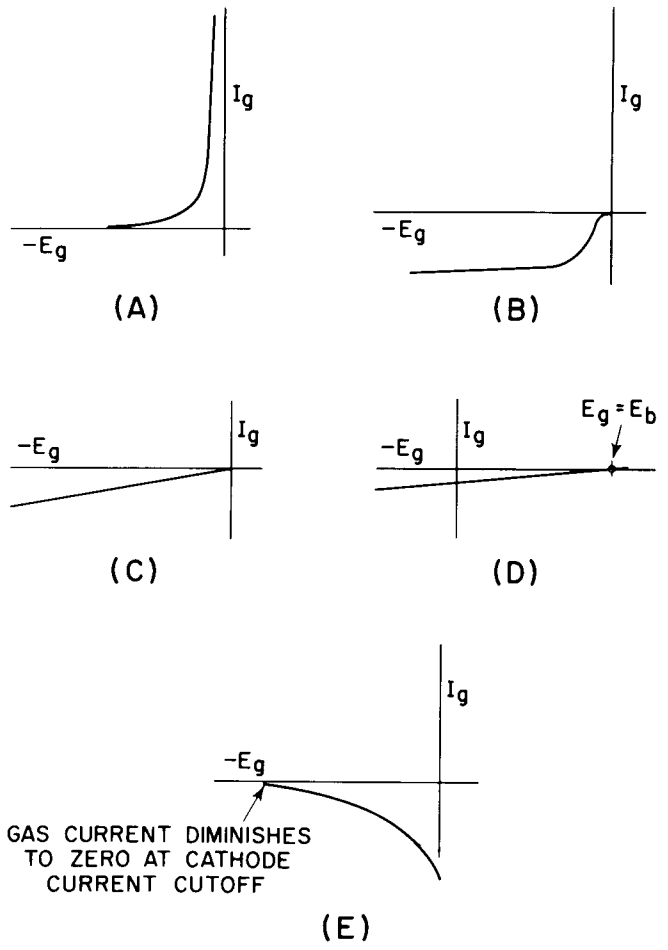


Figure 7. Individual Currents which Contribute to the Composite Grid-Characteristic Curve of Fig. 8: (A) cathode-to-grid emission, (B) primary grid emission, (C) grid-to-cathode leakage current, (D) grid-to-plate leakage current, (E) gas current

small, but definite, probability that an electron leaving the cathode will collide with a gas molecule and form a positively charged ion. Ions so formed move toward the most negative electrode (usually the grid), and become neutralized by electrons from the grid surface. In effect, then, electrons are leaving the grid in the same direction as they do for grid emission and for leakage currents. The current resulting from this flow of electrons to neutralize ions is termed gas current, and its characteristic curve is shown in Fig. 7E.

When the grid bias voltage is large, the gas current is zero because the cathode current is cut off and, therefore, no ions can be formed. As the grid bias voltage is reduced and the plate current increases, the number of ions which are formed increases; the ions continue to go to the grid as long as it is the most negative element within the tube. In a description of the curves of Fig. 7E there is a complicating factor which comes from the probable increase of gas evolution rate with the increase in plate temperature caused by the higher plate current. Thus, the shape of the curve will depend upon the gas-evolution characteristics of the plate and the opposing gas-adsorption characteristics of the cathode

and getter, both effects combining at any instant to determine the gas pressure in the tube.

THE COMPOSITE GRID CHARACTERISTIC

In all receiving tubes, all of the grid currents previously described are present to some degree under operating conditions. An oscillographic display of the grid-current-vs.-grid-voltage characteristics is therefore a composite of the individual curves of Fig. 7; such a composite is shown in Fig. 8. The individual contributions to the total grid current can be determined by observation of the curve shape with and without plate voltage being applied. When there is no plate voltage, only grid-to-cathode leakage electrons and grid-emission electrons subtract from the electrons emitted from the cathode and entering the grid. The currents due to grid emission and to grid-to-cathode leakage can be separated, as shown in Fig. 8, by assuming that the constant component of grid current below the axis is due to grid emission, and the sloping component is due to leakage current. The leakage resistance can be calculated from the slope, or from the leakage current at some arbitrarily chosen reference grid voltage chosen to provide a relative measure of leakage resistance. When plate voltage is applied, grid-plate leakage and gas currents are added to the composite curve. Since the characteristic curve for grid-to-plate leakage current, Fig. 7C, is almost a level line in the relatively short range of grid bias voltage represented in Fig. 8, it can appear almost as an increase in grid emission rather than as an added leakage component. Actually, it is possible that the grid emission will increase upon the application of plate voltage simply because the outer side of each grid wire is exposed to an added electric field. However, experience has indicated this is not as likely as is the presence of grid-to-plate leakage. Once again, the increase in current at some arbitrarily chosen reference grid-bias voltage can be used as a measure of the added grid-to-plate leakage current.

The magnitude of the gas current varies nonlinearly with grid-bias voltage and, therefore, only a value at some reference point can be used as its measure. The reference point is generally chosen as the value between the lowest portion of the gas curve and the line extrapolated from the bias voltage above which plate current is cut off (i.e., where no gas current can exist). The choice can be made visually from an oscillographic curve trace or, in dc measurements, by means of a plotted curve. As mentioned previously, the length of time that a tube is held at any one condition of bias voltage and plate current can alter the gas current if the holding causes an increase in plate temperature and an increase in gas evolution. A slow point-by-point measurement to obtain a curve can yield a different result than if the curve is obtained quickly.

INTERPRETATION OF DATA

The ability to use information on grid currents and surface potential depends upon the kind and amount of data available because all of the effects discussed so far are interrelated. It will be assumed that the complete grid-current-vs.-grid-voltage and grid-voltage-vs.-plate-current curves are available. Test apparatus

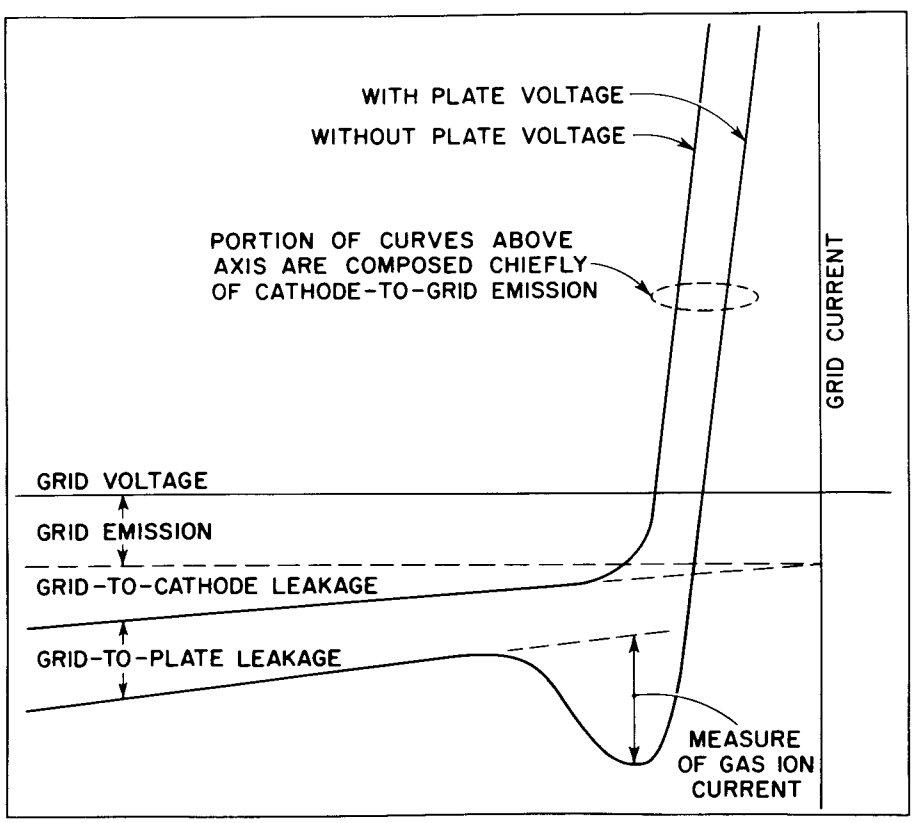


Figure 8. Composite Grid-Characteristic Curve Containing the Individual Currents from Fig. 7

to display both curves simultaneously has been developed and its use promises to become a standard method.

Grid characteristics are used chiefly to assist in the evaluation of tube behavior in the light of changes which occur during life or due to planned variations in materials and processing. Because the shape of the grid curves is affected by many variables, as mentioned previously, there are no clear statements which can be made concerning cause and effect of shifts in the curves. For instance, if the cathode is heated to higher than normal temperatures for some extended time early in tube life, a greater amount of subliming cathode coating will deposit on the grid and mica parts to cause an increase in leakage and, possibly, grid emission, and a decrease in the grid work function. If the same process were to be accompanied by increased gas evolution, however, it is possible that the grid surface would be poisoned by the gas and that the grid work function and grid emission would change in a direction opposite to the previous case. A great deal can be surmised by monitoring the curve shapes before and after each step. Care must be exercised in comparing grid curves between different tubes of the same type, such as would be necessary in evaluating a change in materials. Such a comparison is not as valid as one made on the same individual tube, because in the former case there are added variables such as differences in tube geometry and emission.

One of the most important reasons for using information from grid characteristics curves is to assess the importance of any change in grid characteristics upon the actual tube bias voltage and the resulting plate current. Consider hypothetically the case of bias-voltage changes which result from a processing step where high cathode temperatures and the resultant increase in sublimation cause a lowering of the grid work function and a slight increase in grid emission and leakage (the gas currents remaining the same). This case is illustrated in Fig. 9 where the original state of the tube is given by solid lines and the final state by dashed lines. If the effect of any changes in grid emission and leakage is discounted, the drop in grid work function is indicated by the lateral displacement of the grid characteristic. Ordinarily, this shift would be specified as a change in 0.1 microampere retarding potential, taken without plate voltage. However, the grid curves without plate voltages were omitted to simplify the diagram.

It was explained earlier that a lowering of the grid work function is actually equivalent to making the grid more positive by the same amount. This change does two things. More electrons enter the grid, raising that portion of the grid curve above the voltage axis and giving it the appearance of moving to the left by an amount equal to the work-function change. It also raises the plate current by the amount it would ordinarily be changed due to the same equivalent bias-voltage shift. The shift of the grid curve to the left accompanied by a rise in plate current often causes confusion because, at first glance, it appears as if an increased bias voltage is resulting in a higher plate current. However, in this case, the position of the grid current curve is being used as a measure of grid-work-function shift and has nothing to do with determining the bias voltage. The bias is determined only by the external voltage applied as modified by the grid work function. Because the externally applied bias voltage in this case involves no grid resistors, the load line (slope = 1/R) is vertical.

If, however, the case is considered of a bias voltage which results only from the voltage drop in a grid resistor (i.e., no applied potential), the intersection of the load line with the grid current-voltage curve is the determining factor in setting the grid potential. Here, the final effective grid potential is a combination of (1) the movement in the positive direction due to the work-function drop, and (2) a movement in the negative direction due to the more negative intersection of the load line with the grid characteristic. The end result is a smaller increase in plate current than occurs without the grid resistor.

If, in the case of the assumed shift in work function mentioned, the grid emission, leakage, and gas currents become large, then the total grid characteristic above the voltage axis shifts less to the left and results in still a different bias voltage and plate current. Usually gas current is the only factor of sufficient magnitude to subtract seriously from the net number of electrons going from cathode to grid as appearing in the positive current region.

In the more complicated case, if an applied external bias voltage is combined with the voltage produced by a grid resistor, the relative role of the solid and dashed grid characteristic lines in determining the bias voltage can be reversed, as illustrated by the displaced load line below the voltage axis in Fig. 9. Obviously, no simple rules can be assumed concerning the influence of grid currents in determining tube bias voltage when grid resistors are used.

REFERENCES

1. Hermann, I. G., and P. S. Wagener, The Oxide-Coated Cathode, p. 13, Chapman & Hall Ltd., London, 1951
2. Chaffee, E. L., Theory of Thermionic Vacuum Tubes, p. 56, McGraw-Hill Book Co., Inc., New York, 1933
3. Dow, W. G., Fundamentals of Engineering Electronics, 2nd Ed., p. 218, John Wiley & Sons, Inc., 1952
4. Dow, W. G., (See Ref. 3, p. 253)
5. Nergaard, L. S., "Studies of the Oxide Cathode," RCA Review, Vol. 13, pp. 512, 520, 496, Dec. 1952
6. Hermann, I. G., and P. S. Wagener, (See Ref. 1, p. 192)
7. Hermann, I. G., and P. S. Wagener, (See Ref. 1, p. 187)
8. Dow, W. G., (See Ref. 3, p. 203)
9. Hermann, I. G., and P. S. Wagener, (See Ref. 1, p. 32)
10. Hermann, I. G., and P. S. Wagener, (See Ref. 1, p. 28)
11. Nottingham, W. B., Handbuch der Physik, Vol. XXI, p. 10, 1956
12. Dingwall, A. G. F., "A Rapid Method for Estimation of Grid Cathode Contact Potential," Pro-

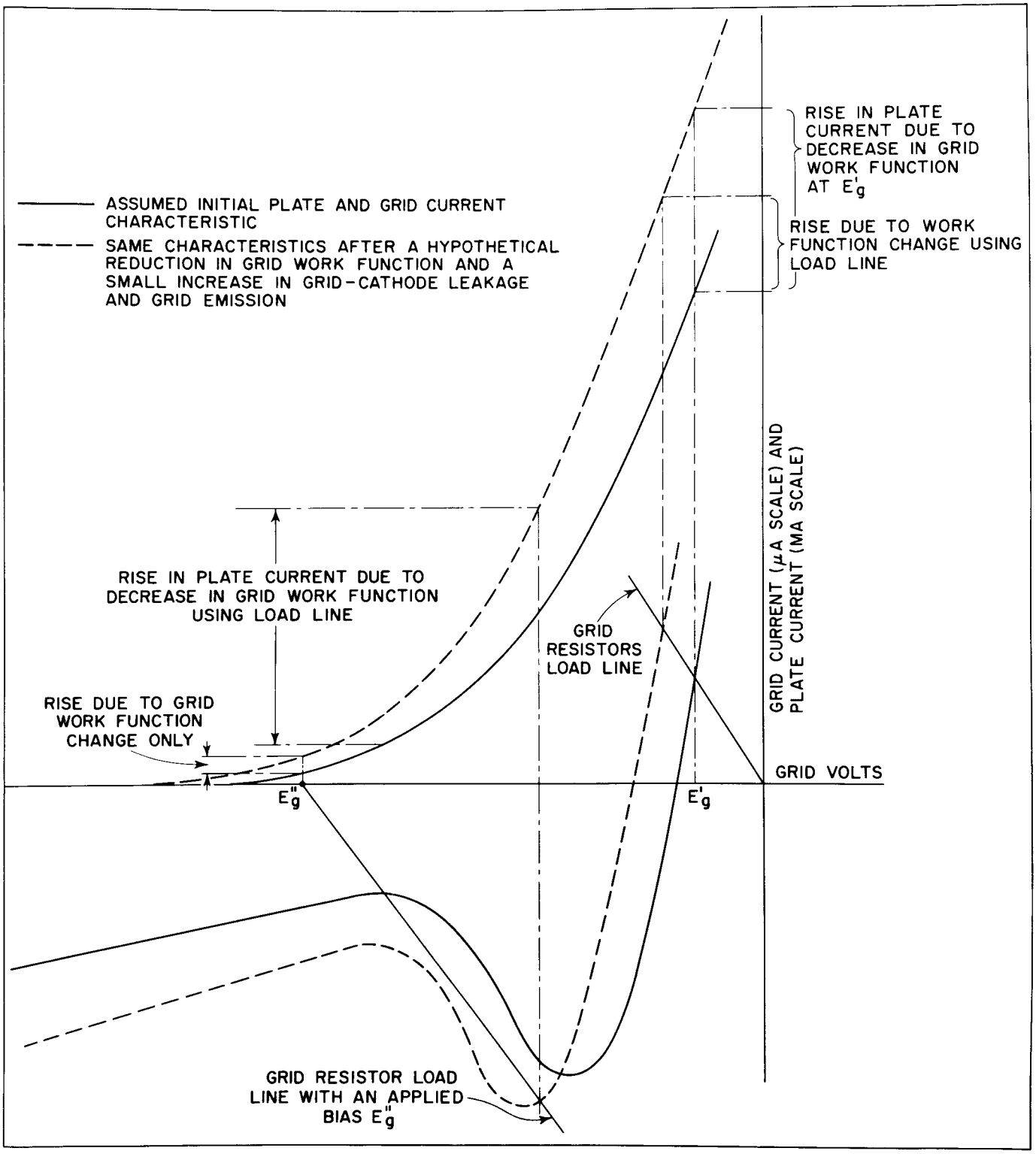

ASSUMED INITIAL PLATE AND GRID CURRENT CHARACTERISTIC
SAME CHARACTERISTICS AFTER A HYPOTHETICAL REDUCTION IN GRID WORK FUNCTION AND A SMALL INCREASE IN GRID-CATHODE LEAKAGE AND GRID EMISSION
RISE IN PLATE CURRENT DUE TO DECREASE IN GRID WORK FUNCTION AT E'g
RISE DUE TO WORK FUNCTION CHANGE USING LOAD LINE
RISE IN PLATE CURRENT DUE TO DECREASE IN GRID WORK FUNCTION USING LOAD LINE
RISE DUE TO GRID WORK FUNCTION CHANGE ONLY
GRID RESISTORS LOAD LINE
GRID CURRENT (μA SCALE) AND PLATE CURRENT (MA SCALE)
GRID VOLTS
E"g
E'g
GRID RESISTOR LOAD LINE WITH AN APPLIED BIAS E"g

Figure 9. Modifications of Plate and Grid Characteristics by a Decrease in Grid Work Function and an Increase in Leakage and Grid-Emission Currents — Hypothetical Case Illustrating Possible Different Effects on Plate Current Changes When Using Combinations of a Grid Resistor and an Applied Bias Voltage

ceedings of the Fifth National Conf. on Tube Techniques, Pergamon Press, 1961

13. Schrader, E. R., "A Survey of Methods Used to Determine Contact Potentials in Receiving Tubes," RCA Review, Vol. XVIII, No. 2, p. 243, June 1957

A New Approach to the Calculation of Electron-Tube Characteristics

O. H. Schade, Sr.

Harrison

Analytic solutions for the space-current flow in electron tubes are known only for simple two-element structures such as the "perfect" parallel-plane diode. The first part of this article shows that the classical method of treating more complex electrode configurations is to reduce them to equivalent diodes by evaluation of an effective mean potential in the plane of the electrode nearest to the cathode. This effective mean potential is determined by averaging the potential field components penetrating into this plane from all electrodes, and summing these averages. The space-current calculation is thereby reduced to the known case. Discussion of this basic theory points out that simplifying assumptions made in the derivation of equivalent-diode potentials become the source of large errors in calculations of modern close-spaced electron tubes.

It is shown in the second part of this article that a more refined method deriving equivalent-diode potentials directly from the electrostatic potential field substantially eliminates the causes for error. For use as a practical engineering tool, a computer program has been set up to carry out the necessarily large mass of detailed calculations by machine as discussed and illustrated in the concluding section of this article.

CLASSICAL METHOD OF ELECTRON-TUBE CALCULATION

THE PROBLEM OF ELECTRON-TUBE CALCULATION

A "perfect" electron tube has no physical or chemical defects. To use the optical term, it has no aberrations and performs precisely as predicted by theory. It is a tube which can be computed precisely because it has the following properties:

1. The cathode has a uniform temperature and work function, and negligible internal resistance.

2. The grids are electron-transparent electrodes that produce homogeneous potential surfaces that coincide with the grid surfaces (have infinitesimal thickness).

3. Electrode spacings are mathematically accurate and have zero deviation at all temperatures and shock conditions.

4. Electrode surfaces have a uniform work function that does not vary with time or temperature.

5. Contacts and connections are ohmic and have negligible resistance.

6. Insulation and environment approach that of a perfect vacuum and remain in stable equilibrium as required by cathode life.

7. Temperatures of cathode and electrodes are constant at predetermined values.

Because of the large number of parameters involved, an infinite number of "perfect" tubes can be designed by varying area, electrode number, electrode spacings, and μ values.

Practical tubes have aberrations, i. e., the preceding conditions can only be approached; the parameters are rarely constants and may even be unstable, so that measured values are, strictly speaking, mean values of functions which are often unknown. This variation is particularly true for close-spaced structures operating at effective potentials considerably smaller than the work functions of metals. The value of calculating the characteristics of a "perfect" tube lies in the numerical relations obtained by varying the parameters. These numerical relations provide theoretical limits and guides for estimating tolerances and operating conditions.

There is little basically new since Langmuir's work, which has been extended or reformulated by various authors[1, 2, 3]. Much work has been done on the problem of calculating the performance of electron tubes having practical grids. The approaches, however, are still very much limited by simplifying assumptions concerning the potential fields and current distribution in the space-charge-filled grid-cathode space. The observation of discrepancies in potential and work-function values, for example, indicates that two potentials exist on the surface of grid wires, one at the front and one at the back, forming a dipole field. The difference in potential is caused by the reduction of the work function of the surface facing the cathode by deposition of evaporated cathode material. This dipole field further complicates matters, underlines nonuniformity, and points up the fact that many measured parameter values must be treated with caution because they represent averages which may depend on the method of measurement.

During the planning of the nuvistors, theoretical

work done to predict their characteristics and behavior included a new look at measurement techniques capable of separating tube parameters. A review of the classic theory of space-current flow in the perfect parallel-plane diode is essential because all practical approaches for calculating the voltage-current characteristics of more complex electron tubes lead to the evaluation of effective potentials for an "equivalent" diode for which the space-current solution is known.

CHARACTERISTICS OF THE "PERFECT" DIODE

Work Functions (ψ), Electron Velocities (V_T), and Saturation Current (I_S)

Metallic conductors contain a very large number of free electrons moving at random in the molecular lattice structure of the metal. Velocities of the electrons follow a probability law. At room temperature nearly all of the free electrons are prevented from leaving the metal by a strong potential barrier, a dipole field formed by the unbalance of molecular or electric forces (image field) at the boundary (surface) of the metal. This potential barrier extends approximately two interatomic distances into space and its value is known as the work function ψ_o of the metal. The work function expressed in volts is a measure of the electron-volt velocity required for the escape (emission) of electrons into a surrounding vacuum in the absence of an external field. Pure metals used in the structure of electron tubes have a work function of about 4.5 volts. Work functions of 2.5 to 1.4 volt can be obtained from alkali metals, or by the addition of monatomic layers of different materials which form complex dipole fields as in the oxide-coated cathodes of electron tubes.

External potential fields that attract electrons can reduce the work function ψ_o. Because of the high negative gradient dv/dx of the potential barrier ψ_o, the potential gradient $E_e = dv_e/dx$ of the external field must be highly positive to reduce ψ_o appreciably, as shown by the following expression:

$$\psi_e = \psi_o - (3.78 \times 10^{-5}\sqrt{E_e}) \qquad (1)$$

The gradient E_e (volts per centimeter) must, for example, have a value $E_e = (0.01 \times 10^5/3.78)^2 = 70{,}000$ volts per centimeter to reduce ψ_o by as little as 0.01 volt. The effect is therefore of little importance in electron tubes that operate with relatively low potential gradients. It is, however, observable when saturated emission currents I_s are measured (see following text) by the fact that I_s does not become independent of potential, but continues to increase slowly with potential.

When the work function has been decreased by the addition of atomic films (as in oxide-coated cathodes), it is not necessarily uniform over the surface area but depends on the surface roughness. Differences in work function between surface elements result in high local gradients which influence the electron-velocity distribution and alter the effect of external fields [Eq. (1)].

As a result, it is necessary to define an effective mean value $\overline{\psi}$ for an emitter, obtained from the solution of the emission equation (see below). This mean work function $\overline{\psi}$ has a temperature dependence of the order of 10^{-4} volt per degree.

In the following discussions, the plain symbol ψ is used to designate the effective work function of an electrode surface. The reader should keep in mind that the effective value ψ of a nonemitting electrode is an average potential value and is not derived from an average emission current.

In comparison with the potential barrier of the work function, the mean volt velocity V_T of the free electrons outside the boundary of the metal is relatively small even at elevated temperatures:

$$V_T = KT/e = T/11605 \qquad (2)$$

A potential barrier $\psi_o = 1.4$ volt for example, is still 14 times higher than the mean electron velocity V_T at 1100 K. To leave the metal and pass the barrier ψ_o, an electron must have an escape velocity larger than ψ_o volts. The probability that an electron has this velocity in a Maxwellian distribution with the mean value V_T is $e^{-\psi_o/V_T}$. For T = 1160 K and $\psi_o = 1.4$, $e^{-\psi_o/V_T} \cong 0.83 \times 10^{-6}$, i.e., about one in 10^6 electrons can escape into space, The number of available free electrons n_o per square centimeter of surface, however, is extremely large:*

$$n_o = 7.5\ T^2 \times 10^{20} \quad \text{cm}^{-2}\ \text{sec}^{-1}$$

Therefore, even this very small fraction is still a large number of electrons. The number of electrons n having an escape velocity larger than ψ_o volts is given by

$$n = n_o e^{-\psi_o/V_T} = 7.5\ T^2 \times 10^{20} e^{-\psi_o/V_T}\ \text{cm}^{-2}\ \text{sec}^{-1}$$

If the electron charge $q_e = 1.6 \times 10^{-19}$ coulombs, the corresponding current density I_s in amperes per square centimeter is given by

$$I_s = nq_e = 120\ T^2\ e^{-\psi_o/V_T} \qquad (3)$$

Substitution of the value for V_T from Eq. (2) provides the following expression for I_s:

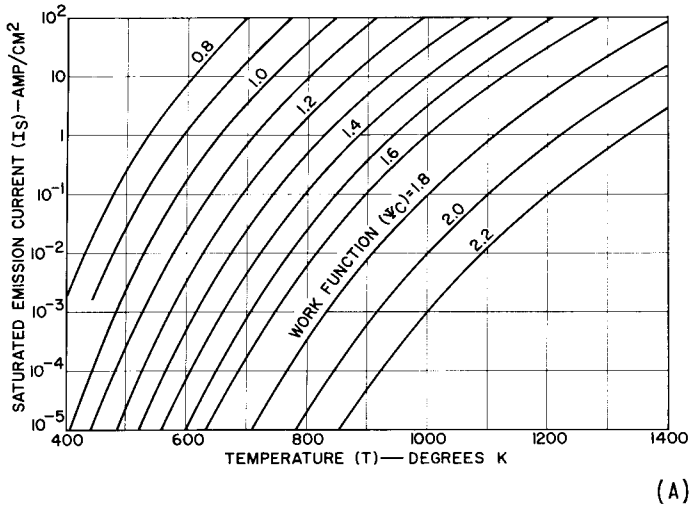
$$I_s = 120\ T^2 \times e^{-\psi_o 11605/T} \qquad (3a)$$

Eq. (3a) is the well known Richardson-Dushman expression for the saturated emission current I_s. This expression is plotted in Fig. 1. The exponential term expresses the fraction of the total number of available electrons which can overcome the potential barrier ψ_o. For the rather low work function $\psi_o = 1.4$ and T = 1160 K, for example, this fraction is $e^{-14} = 0.83 \times 10^{-6}$; therefore, $I_s = 161 \times 10^6\ (0.83 \times 10^{-6}) \cong 133$ amperes per square centimeter.

Control grids and collecting electrodes in an electron tube should have negligible electron emission. It would seem that the high work function $\psi_o \cong 4.5$ of the metals (Ni, Fe, Cu, Mo) used for these electrodes permits a

*The number n_o of available electrons given is obtained from Eq. (3) by letting $\psi = o$ and dividing by the electron charge.

relatively high operating temperature without noticeable emission. In tubes with oxide-coated cathodes, however, it is practically impossible to prevent deposition of cathode material on surfaces visible to the cathode during aging or subsequent operation of the cathode. The work function ψ_o of the "contaminated" portion of the electrode surfaces is thereby often reduced to values approaching the work function of the cathode. For a grid current of 0.1 microampere, a (nuvistor) control grid having a contaminated surface area of 0.06 square centimeter would permit a saturated grid-emission current I_s=1.67 microamperes per square centimeter. Fig. 1B shows that a grid work function $\psi_{o(g)}$ = 1.4 (that of a good cathode) limits the grid temperature to 525 K; a work function $\psi_{o(g)}$ = 1.6 permits a temperature of 600 K. The thermal design of the electrode structure for conservation of heat in the cathode and removal of heat from the remaining electrodes is, therefore, an essential part of electron-tube design.



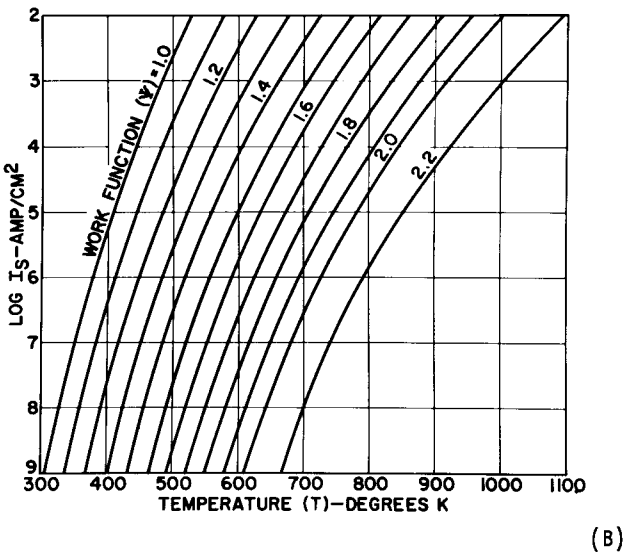


Figure 1. Saturated Emission Current I_s as a Function of Emitter Temperature and Work Function [Eq. (3a)] : (A) temperature=400 K to 1400 K; (B) temperature = 300 K to 1100 K

The energy expended by the electrons in the plate current I in overcoming the potential barriers at the emitter is supplied by the thermal energy of the cathode. The cooling effect or heat loss P_e is given by

$$P_e = I(\psi_c + V_m) + 2V_T \text{ watts} \qquad (4)$$

where V_m is the barrier potential. This cooling effect, which reduces the cathode temperature, becomes appreciable at large currents and affects I_s and I. The potential barrier V_m is caused by the electron charges in space, as discussed below.

The Diode Circuit (Potentials and Operating Conditions)

A diode is basically a two-element capacitor in which the dielectric charge Q due to an applied potential V_d can be moved continuously in one direction by making the negative electrode a good emitter (cathode) and the positive electrode a nonemissive collector (anode). The current I in such a diode system is given in first approximation by the total charge $Q = CV_d$ divided by the transit time τ required to move the charge through the distance d (centimeters) of the dielectric space between electrodes. A thermionic diode has a vacuum dielectric for which the dielectric constant is unity. The charge density Q/A of a parallel-plane vacuum capacitor is, therefore, $Q/A = CV_d/A = (10^{-11}/36\,\pi\,d)V_d$ coulombs per square centimeter. The electron transit time τ is the distance divided by the mean velocity:

$$\tau = d/\bar{v} = d/(1/2)\ (2V_d \frac{e}{m})^{1/2}$$

$$\tau = 2d/5.93 \times 10^7 V_d^{1/2}$$

The current density I for a parallel-plane diode is, therefore, in first approximation:

$$I = Q/A\ \tau = \frac{5.93 \times 10^{-4}}{72\,\pi\,d^2} V_d^{3/2}$$

$$I = (2.63 \times 10^{-6}/d^2)\ V_d^{3/2} \qquad (5)$$

The potential V_d is the potential difference between the barrier potentials ψ_c and ψ_a of the electrodes (see Fig. 2B). Except for an 11-per-cent smaller constant, Eq. (5) is the well known 3/2-power-law relation (Child's law) of the thermionic parallel-plane diode. The relation is a first-order approximation because it assumes an unlimited supply of electrons with zero initial velocity and neglects the effect of electron charges on the potential distribution in the dielectric (vacuum). Reasonably good agreement between observed and computed values can therefore be expected only for currents $I \ll I_s$ and potentials $V_d > 10$ volts, which are large compared to initial electron velocities. In modern close-spaced diodes ($d \simeq 0.005$ centimeter), the initial electron velocities alone are sufficient to promote a substantial current even when the potential V_d is negative. These considerations indicate that there are three significantly different potential distributions in the vacuum dielectric of the diode. The distribution depends on the magnitudes of the external anode potential E_b, the electron-volt potentials V_T, the space

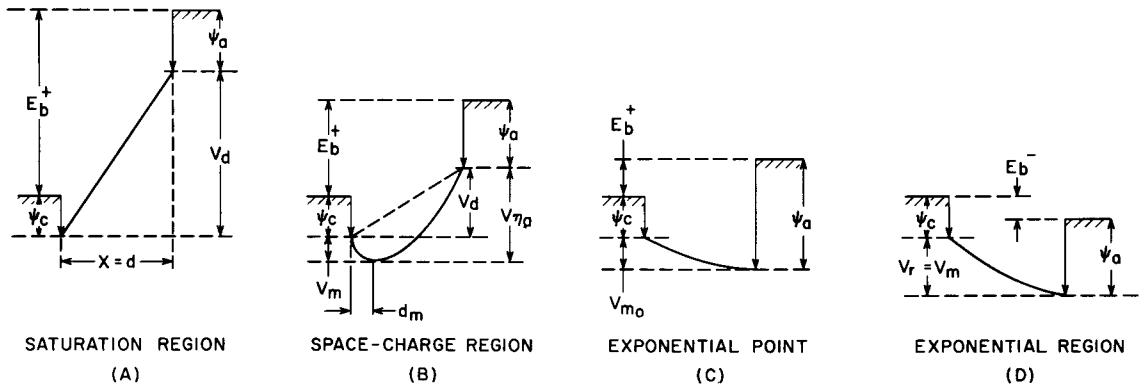


Figure 2. Operating Regions of a Diode: (A) saturation region; (B) space-charge region; (C) exponential point; (D) exponential region

charge of the electrons emitted into space (because of the thermal energy supplied to the cathode), and the electrode barrier potentials ψ_c and ψ_a, as illustrated in Fig. 2.

The Saturation Region (Fig. 2A). The external potentials E_b and V_d are so high that all electrons that have an initial volt-velocity greater than the work function barrier ψ_c of the cathode are rapidly accelerated towards the anode. The linear electrostatic potential distribution in the vacuum space is essentially undisturbed by the low electron-charge density in space. The potential gradient dv/dx is positive and substantially constant; a potential minimum does not exist. The saturation current $I = I_s$ is given by Eq. (3). For very high values of E_b, ψ_o in Eq. (3) may be replaced by the effective value ψ_e of Eq. (1) to include the reduction of the work-function barrier by a high-field gradient in the vacuum dielectric.

The Space-Charge-Limited Region (Fig. 2B). The potentials E_b and V_d are positive, but insufficient to move all electrons having passed ψ_c to the anode. Because of the increased transit time τ, the charge density in space is increased sufficiently to depress the space potential to a value below that of the electrostatic potential (broken line) and to cause the formation of a space-charge barrier near the cathode. This space-charge barrier is the "minimum potential" V_m. The minimum potential occurs at a distance $x = d_m$ which is passed only by electrons having a volt-velocity greater than $(\psi_c + V_m)$. The field gradient dv/dx is therefore negative for distances $x < d_m$ and positive for distances $x > d_m$. The accelerating field has a potential $V\eta a$ larger than V_d, i.e., $V\eta a = V_d + V_m$. The initial volt-velocity of the electrons tends, therefore, to increase the space current I, while the space charge (V_m) reduces I by blocking low-energy electrons. "Langmuir's solution" for the space-charge-limited current, discussed below, contains 6 parameters (the potentials V_m and V_η, the distances d and d_m, the cathode temperature T, and the work function ψ_c) as compared to the two parameters (V_d and d) of the rough approximation given by Eq. (5).

The Exponential Point (Fig. 2C). When the diode potential V_d is made slightly negative, the minimum potential V_m increases and moves toward the anode. The exponential point represents the condition where the minimum V_m occurs at the anode surface ($d_m = d$) and the accelerating potential $V\eta a$ equals zero. The current has a particular value designated by I_o.

The Exponential Region (Fig. 2D). In the exponential region, the electrons are retarded as they pass through the entire distance d; the potential gradient at the anode is negative and the minimum has disappeared. Only electrons having a velocity greater than the total barrier $(\psi_c + V_d)$ can reach the anode. Using the notation used to derive Eq. (3), the number of electrons n is given by the small fraction $n = n_o e^{-(\psi_c + V_d)}$. The corresponding current I_r is given by

$$I_r = 120\ T^2\ e^{-(\psi_c + V_d)/V_T}$$

Using Eq. (3a), $I_r = I_s e^{-V_d (11605/T)}$

$$\text{where } V_d = V_m < V_{m_o} \qquad (6)$$

Calculation of the limit V_{m_o} set by the exponential point requires the use of the Langmuir relations explained in the following section. It is apparent that the current I_o can then be computed by Eq. (6) with $V_d = V_{m_o}$.

Effective Anode Potential V and Reference Condition (V = 0). For calculation of the characteristic $I = f(V)$, the exponential point (Fig. 2C) is a preferred reference for defining an "effective" anode potential V. This point indicates the condition dv/dx = 0 at the anode which is not indicated by the potential V_d. The potential reference point V = 0 is thus defined by

$$V = 0 \text{ for } dv/dx = 0,\ V\eta a = 0 \text{ and } I = I_o \qquad (7)$$

where I_o is the reference current value at $V = V\eta a = 0$. The reference condition defined by Eq. (7) facilitates the calculation of equivalent diode potentials from the electrostatic fields of practical triodes discussed in the second part of this article. The relation of the effective potential V to the potentials in the diode circuit is illustrated in Fig. 3. The reference condition (exponen-

tial point) V = 0, I = I_o is shown in Fig. 3A. The total barrier potential V_o at the cathode which must be overcome by the electrons in the reference current I_o is

$$V_o = (\psi_c + V_{mo}) \tag{8}$$

For accelerating fields (Fig. 3B), $V_{\eta a} > 0$, $I > I_o$, the effective potential is positive:

$$V^+ = V_{\eta a} + (V_{mo} - V_m) \tag{9}$$

For the exponential region (Fig. 3C), $V_{\eta a}$ does not exist, $I < I_o$, and the effective potential V is negative:

$$V^- = (V_{m_o} - V_m) \tag{10}$$

The external anode voltage E_b is given by the following:

$$E_b = V - V_o + \psi_a = V - (\psi_c + V_{m_o}) + \psi_a \tag{11}$$

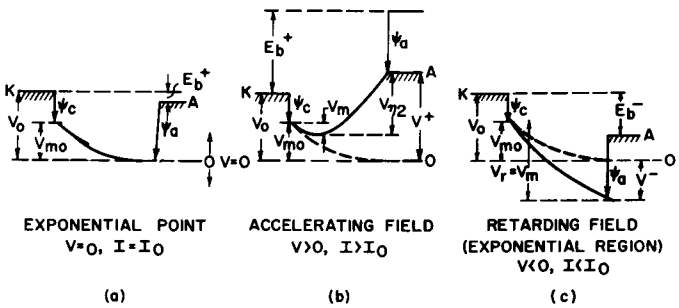


Figure 3. Effective Diode Potential $V^{\pm}$: (A) reference condition V = 0, exponential point dv/dx = 0, I = I_o; (B) accelerating fields V > 0, dv/dx positive, I > I_o; (C) retarding fields V < 0, dv/dx negative, I < I_o

The Langmuir Relations

Langmuir has expressed the potential distribution v = f(x) in the vacuum dielectric of a parallel-plane diode by a single normalized function $\eta = f(\xi)$ by introducing dimensionless distances ξ and potentials η. This function is illustrated in Fig. 4. The normalized distances ξ are measured from the location of the potential minimum, where $\xi = 0$ and $\eta = 0$. The actual cathode-anode distance d is, therefore, the sum of two normalized distances:

$$d = l (\xi_c + \xi_a) \tag{12a}$$

The cathode-to-potential-minimum distance d_m is given by

$$d_m = l \xi_c \tag{12b}$$

The potential-minimum-to-anode distance $(d - d_m)$ is given by

$$(d-d_m) = l \xi_a \tag{12c}$$

The normalizing factor l is given by

$$l = \frac{(T/1000)^{3/4}}{13.1\ I^{1/2}} \tag{12d}$$

for I in amperes and distances d in milli-inches (mils).

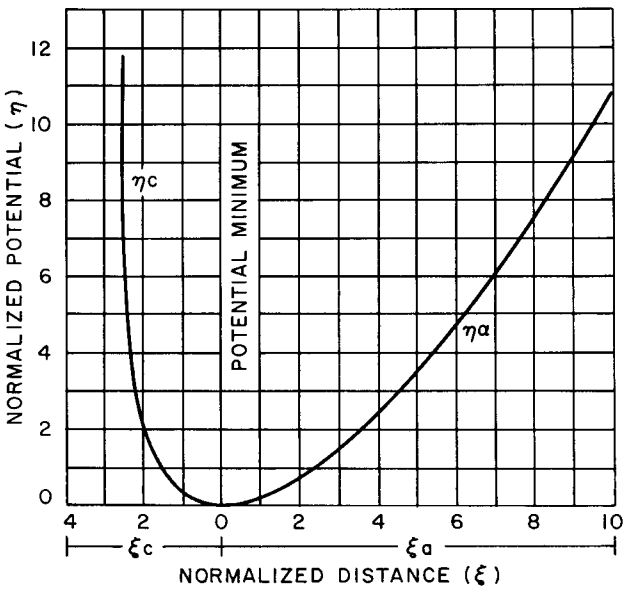


Figure 4. Langmuir Function $\eta = f(\xi)$ for the Space-Charge-Limited Region V > 0

The normalized potential η between cathode surface and potential minimum depends on the space-charge density and is, therefore, a function of the ratio of the anode current I to the saturated emission current I_s:

$$\eta_c = \ln(I_s/I) \tag{13}$$

The current I_s is given by Eq. (3). The associated anode potential η_a with respect to the potential minimum is determined by the distance relation, Eq. (12a), as follows:

With $\xi_c = f(\eta_c)$ from the cathode branch of the Langmuir function (Fig. 4), or $\xi_c = f(I_s/I)$ directly from the graph shown in Fig. 5, calculate ξ_a from Eq. (12a), i. e., $\xi_a = (d/l) - \xi_c$. The results then determine the value $\eta_a = f(\xi_a)$ in the anode branch of the Langmuir function shown in Fig. 4 (or more accurately in Fig. 6). The corresponding potentials $V_{\eta c} = V_m$ and $V_{\eta a}$ measured in volts are obtained by multiplication by the normalizing factor $V_T = (T/11605)$ from Eq. (2):

$$V_\eta = \eta_c (T/11605)$$

Then by substitution of Eq. (13):

$$V_\eta = 0.198\ (T/1000) \log I_s/I \tag{14}$$

Similarly, $V_{\eta a}$ is given by

$$V_{\eta a} = \eta_a (T/11605) \tag{15}$$

Calculation of the Characteristic I = f(V)

Most quantities in the Langmuir relations are functions of the current. The (I, V) characteristic is, therefore, more readily obtained by making the current the independent variable. After the parameters T, ψ_c,

and the diode spacing d are selected, the following procedure is used:

1. Determine the reference current I_o at the exponential point V = 0. An expression for I_o is obtained from Eq. (12b) for a value of d_m = d:

$$I_o = 5.83\ \xi_c^2\ (T/1000)^{3/2}/d^2 \times 10^3\ \text{Amp.} \qquad (16)$$

where d = diode spacing in mils. Eq. (16) can be solved directly for $I_s = \infty$, with the asymptotic value $\xi_c = 2.554$.* For finite values I_s/I_o, ξ_c and I_o have lower values and Eq. (16) must be solved by successive approximation and interpolation, as shown in Table I for d = 2 mils. Assume logarithmically spaced arbitrary values I_s/I_o. Tabulate the corresponding values ξ_c obtained from Fig. 5 (obtained from Eq. (2) and the Langmuir function). Tabulate I_o computed from Eq. (16) for a given T, and the values $I_s = I_o \times (I_s/I_o)$. The function $I_o = f(I_s)$ can then be plotted, as shown in Fig. 7.

2. Select a cathode work function ψ_c and an emission current I_s at the desired temperature T from Fig. 1A. List I_o as the second value of the independent variable, as shown in Table II (first column), and compute I_s/I for the second column.

3. List $\xi_c = f(I_s/I)$ obtained from Fig. 5.

4. Compute the normalized cathode-anode distance $(\xi_c + \xi_a)$ from Eq. (12a)

5. Calculate $\xi_a = (\xi_c + \xi_a) - \xi_c$

6. Obtain $\eta_a = f(\xi_a)$ from Fig. 6 (Langmuir table)

7. List the values V_m obtained from Fig. 8 (or Eq. 14) for the ratios I_s/I.

8. Calculate V^+ or V^- from Eq. (9) or (10)

The characteristics of some "perfect" diodes are plotted in Figs. 9, 10, and 11.

Calculation of the Characteristic I = f(E_b)

If the effective potential V has been computed as outlined in the preceding section, the external plate voltage is obtained by means of Eq. (11), in which all quantities are known except the work function ψ_a of the anode. The work function of pure metals used as plate materials is generally too high in diodes having oxide cathodes because of cathode-coating material evaporated during exhaust; aging reduces ψ_a to about 2.5 to 3 volts.

It may be more convenient for some purposes to eliminate the reference value V by combining Eqs. (9) and (11), as follows:

$$E_a = (V\eta_a - V_m) + (\psi_a - \psi_c) \qquad (17)$$

The following steps are then required:†

1. Select the parameters T, ψ_c, and d, and determine I_s from Fig. 1.

2. Assume a value for the current I, and determine V_m from Fig. 8.

3. Calculate I_s/I, and obtain $\xi_c = f(I_s/I)$ from Fig. 5.

4. Calculate ξ_a by means of Eq. (12a).

5. Obtain $\eta_a = f(\xi_a)$ from Fig. 7.

6. Calculate $V\eta_a$ by means of Eq. (15).

7. Calculate E_b by means of Eq. (17).

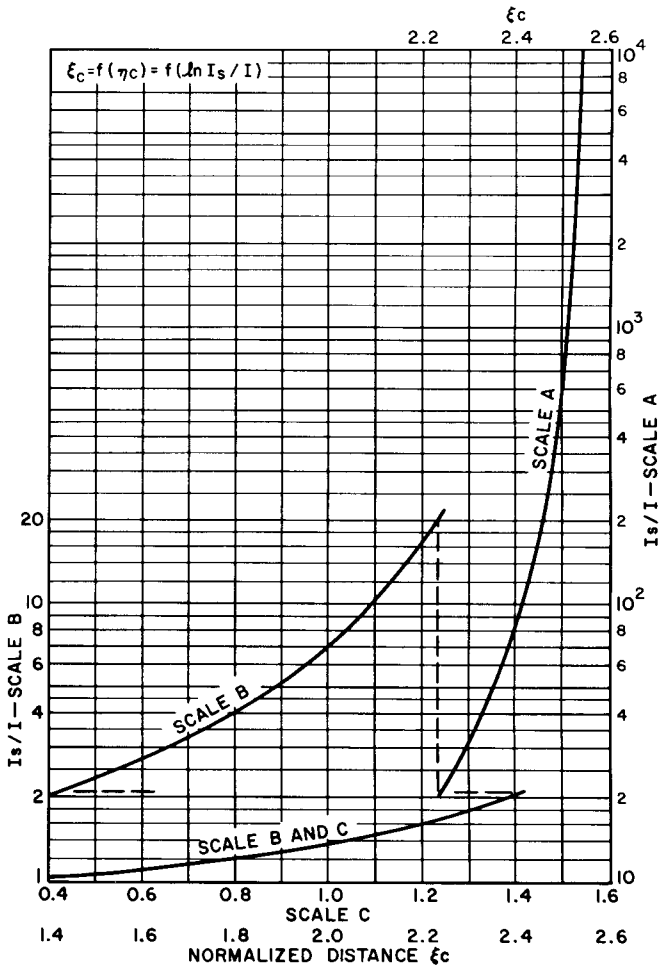


Figure 5. The Function $\xi_c = (I_s/I)$

Some Properties of the Perfect Diode

The barrier potential $(\psi_c + V_m)$ which must be overcome by the electrons in the anode current I can be calculated by solving Eq. (3) for ψ_c and adding the expression for V_m from Eq. (14):

$$(\psi_c + V_m) = 0.198(T/1000) \log 1.2(T/1000)^2\ 10^8/I \qquad (18)$$

This value is independent of the work function ψ_c of the cathode when T and I are constant. In a normal diode with fixed spacing d, however, the current is not independent of ψ_c, as shown by Eq. (16). The change

*The current I_o calculated for $I_s = \infty$ is the value I_∞ in Ref. 3.
†Method used in Ref. 4.

$\Delta\ (\psi_c + V_m)$ is relatively small, but the change ΔI is appreciable in close-spaced high-conductance diodes, as shown by Table III.

The changes of barrier potential and current are noticeable even for current ratios $I_s/I > 100$, a value 10 times larger than that quoted in the literature.

Table I

Calculation of $I_o = f(I_s)$ for d = 2 milli-inches

$I_o = 1.448\ \xi_c^2 (T/1000)^{3/2}$ ma							
		T = 900 K		T = 1000 K		T = 1100 K	
I_s/I_o	o	I_o	I_s	I_o	I_s	I_o	I_s
6570	2.536	7.97	52.5	9.35	61.5	10.80	71.0
2820	2.528	7.90	22.3	9.27	26.1	10.70	30.2
666	2.50	7.74	5.15	9.06	6.04	10.46	6.97
298	2.471	7.52	2.24	8.84	2.63	10.20	3.04
117.5	2.42	7.25	0.855	8.50	1.00	9.81	1.155
47.3	2.35	6.81	0.322	8.00	0.378	9.24	0.436
16.65	2.20	5.99	0.094	7.02	0.110	8.10	0.127

Table II

Computation of Diode Characteristics for d = 2 milli-inches

I	I_s/I	ξ_c	$(\xi_c+\xi_a)$	ξ_a	η_a	V_m	$V = (V_{m_o} V_m) + \eta_a (T/11605)(\xi_c+\xi_a)\frac{T}{11605} g$			
T = 1100 K, I_s = 7.0 A, ψ_c = 1.6 V										
0.0035	2000					0.72	-0.104	-0.104	$24.4\sqrt{I}$	0.095
0.0104	670			0.0	0.0	0.616	0	0		
0.020	350	2.48	3.45	0.97	0.203	0.555	0.061 + 0.0193 = 0.080			
0.040	175	2.448	4.88	2.432	1.06	0.488	0.128 + 0.1005 = 0.2285			146.
0.050	140	2.434	5.46	3.026	1.56	0.467	0.149 + 0.148 = 0.297			
0.080	87.5	2.4	6.9	4.5	3.02	0.421	0.195 + 0.287 = 0.482			
T = 1050 K, I_s = 2.8 A, ψ_c = 1.6 V										
0.03	934					0.618	-0.105	-0.105	$25.3\sqrt{I}$	0.0905
0.0095	295			0.0	0.0	0.513	0	0		
0.020	140	2.433	3.58	1.147	0.27	0.446	0.067 + 0.0244 = 0.0914			
0.040	70	2.382	5.06	2.678	1.25	0.384	0.129 + 0.113 = 0.242			143.
0.050	56	2.362	5.66	3.298	1.78	0.362	0.151 + 0.161 = 0.312			
0.080	35	2.312	7.15	4.838	3.44	0.32	0.193 + 0.311 = 0.504			
T = 1000 K, I = 1 A, ψ_c = 1.6 V										
0.0033	300					0.491	-0.081	-0.081	$26.2\sqrt{I}$	0.0861
0.0085	117.5	2.42		0.0	0.0	0.410	0	0		
0.020	50	2.35	3.70	1.35	0.37	0.336	0.074 + 0.0319 = 0.106			
0.040	25	2.268	5.24	2.972	1.49	0.276	0.134 + 0.128 = 0.262			
0.050	20	2.232	5.86	3.628	2.1	0.257	0.153 + 0.181 = 0.334			139.
0.080	12.5	2.145	7.40	5.255	3.92	0.216	0.194 + 0.338 = 0.532			
T = 950 K, I_s = 0.35 A, ψ_c = 1.6 V										
0.003	116.5					0.388	-0.075	-0.075	$27.7\sqrt{I}$	0.0818
0.0074	47.3			0.0	0.0	0.313	0	0		
0.020	17.5	2.21	3.92	1.71	0.56	0.232	0.081 + 0.0457 = 0.127			
0.040	8.75	2.055	5.55	3.495	1.97	0.175	0.138 + 0.161 = 0.299			
0.050	7.0	2.00	6.20	4.20	2.70	0.157	0.156 + 0.221 = 0.377			128.
0.080	4.37	1.83	7.84	6.01	4.85	0.117	0.196 + 0.396 = 0.592			

Table II (continued)

Computation of Diode Characteristics for d = 2 milli-inches

T = 900 K, I_s = 0.1 A, ψ_c = 1.6 V								
I	I_s/I	ξ_c	$(\xi_c+\xi_a)$	ξ_a	η_a	V_m	$(V_{m_o} V_m) \overset{V=}{+} \eta_a (T/11605)(\xi_c+\xi_a)\frac{T}{11605}g$	
0.002	50					0.300	-0.086 -0.086 28.4 $\sqrt{I}$ 0.0775	-
0.00602	16.1			0.0	0.0	0.214	0 0	
0.020	5.0	1.885	4.02	2.135	0.84	0.122	0.092 + 0.065 = 0.157	
0.040	2.5	1.54	5.69	4.15	2.64	0.068	0.146 + 0.2025 = 0.3485	108.
0.050	2.0	1.38	6.35	4.97	3.58	0.050	0.164 + 0.277 = 0.441	
0.080	1.25	0.85	8.03	7.18	6.5	0.016	0.198 + 0.504 = 0.702	
0.100	1.0	0	9.00	9.00	9.25	0	0.214 + 0.716 = 0.930	

Table III

Change of Barrier Potential $\Delta(\psi + V_{m_o})$ and ΔI_o for a Change of Cathode Work Function from ψ_c = 1.4 v to ψ_c = 1.6 v (Example for the Exponential Point I = I_o)

T = 900 K d = 0.002 in.							
ψ_c	I_s	I_o	I_s/I_o	V_{m_o}	$\psi_c + V_{m_o}$	$\Delta(\psi_c + V_{m_o})$	% ΔI_o
1.4	1.32	0.00741	178.0	0.401	1.801	0.016	-19%
1.6	0.1	0.00601	16.65	0.201	1.817		
T = 1000 K d = 0.002 in.							
1.4	10.0	0.00914	1094.0	0.601	2.001	0.01	-7%
1.6	1.0	0.0085	117.6	0.411	2.011		

Exponential Region V < 0. In diodes with A = 0.2 square centimeter and d = 0.002 inch, the exponential region occurs below 2 milliamperes at T = 1100 K and below 1.2 milliamperes at T = 900 K. Note that the slope of the I and g functions increases with decreasing temperature (Fig. 10), but does not change with ψ_c (Fig. 11).

Positive-Accelerating-Potential Region (V > 0). For d = 0.002 inch, the equivalent grid potential V^+ is approximately 0.3 volt for the normal temperature T = 1100 K and a current density of 50 milliamperes per square centimeter. It is obvious from Fig. 3b that V^+ is only a small fraction of the anode work function ψ_a, which is replaced by the grid work function ψ_g in a triode.

Diode Conductance $g = f(T, \psi_c, d)$. In the retarding-field region, V ≦ 0, the conductance is given by

$$g = (11605/T)\ I \text{ milliamperes per volt.} \qquad (19)$$

It has a discontinuity at V = 0 and increases much more slowly in the accelerating range V > 0, where it can be computed by special formulas (see Refs. 2, 3), or simply from calculated incremental values $\Delta I / \Delta V$, or from an accurate graph of the computed characteristic I = f(V). The conductance decreases with cathode temperature. The inset in Fig. 10 shows the reduction of g as a function of T in the perfect diode (d = 0.002 inch) for a constant work function ψ_c = 1.6 volt for two values of constant potential V, V = 0 (computed by means of Eq. (19), and V = 0.262 volt (taken from graph); and for a constant current I = 45 milliamperes per square centimeter. These graphs illustrate the theoretical change of g with heater voltage. In a nuvistor the temperature change is $\Delta T \simeq -60$ K when the heater voltage is reduced from 6.3 to 5.5. The upper portion of Fig. 10 shows that the change Δ_g is reduced by a factor of two when the cathode work function is decreased from ψ_c = 1.6 volt to ψ_c = 1.4 volt.

The deterioration of g with time (life) because of loss of emission (I_s) can be appraised by computing the characteristics of the perfect diode for a constant temperature and varying the cathode work function ψ_c. The results of such a computation for T = 900 K and d = 0.002 inch are shown in Fig. 11. The inset shows g as a function of decreasing cathode work function for constant potential V and constant current. It can be seen that low-temperature operation requires very low work functions. Other causes for a reduction of I and g are the degenerative effects of interface and bulk resistance, which, in effect, represent a resistance in the cathode circuit.

The external plate voltage is the sum $E_b = V + \psi_a - V_o$. For tests made at one time, the anode work function

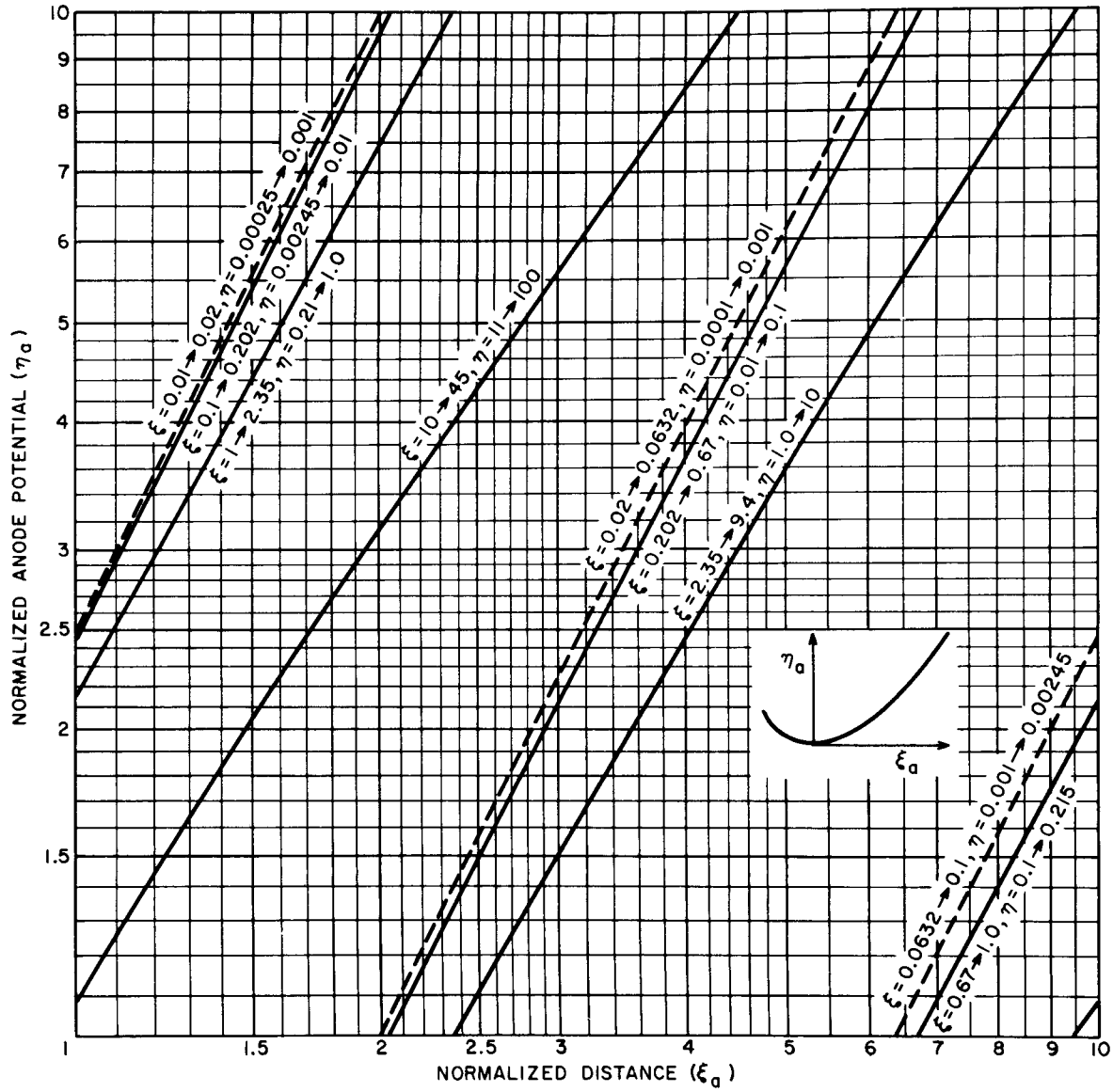


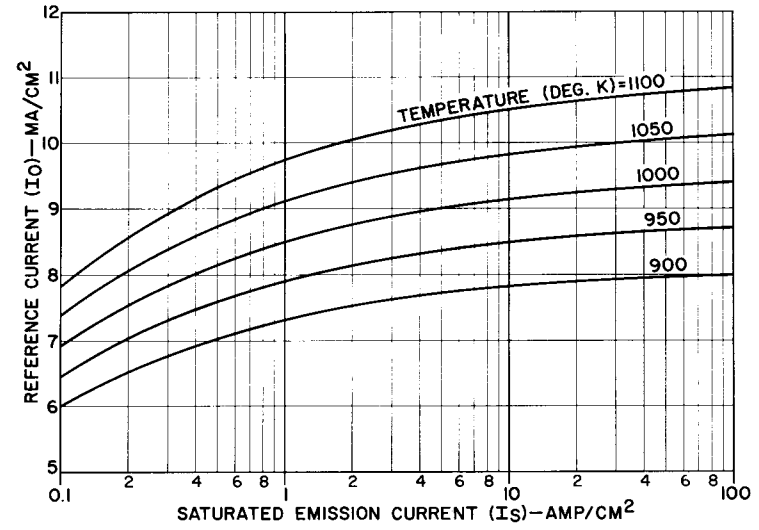
Figure 6. Langmuir Function $\eta_a = f(\xi_a)$

ψ_a can be assumed to be constant. The effective diode potential $V = (E_b - \psi_a) + V_o$ is a function of temperature because V_o varies with T. The table in Fig. 10 shows that the change ΔV_o for $\Delta T = 50$ K is substantially constant at $\Delta V_o \simeq 0.1$ volt.

For fixed-external-voltage tests, therefore, $V = E_b - \psi_a + (V_{o(o)} + \Delta V_o)$. For values of $T_o = 1100$ K and $V_{o(o)} = 2.216$. Table IV can be constructed with the use of Fig. 10.

The increments ΔV_o are indicated in Fig. 9, which shows the result of a similar evaluation for a perfect diode with 0.0025-inch spacing, as well as for a perfect triode ($\mu = 32$), as discussed below. Because of the positive curvature of the characteristics, tests made at a higher current result in smaller ΔI and Δg values than tests at a lower current. Fixed-voltage testing causes much larger current and conductance changes than those obtained for constant effective potential or constant current. Furthermore, the anode work function ψ_a (replaced by the grid work function ψ_g in triodes) can easily be decreased by 0.5 volt or more by barium deposits evaporated from the cathode during processing of a tube. A less negative value of ψ_a increases I and g, and results in smaller changes in ΔI and Δg when the temperature is decreased by a reduction of heater voltage. The apparent improvement has often been attributed to higher emission. This misinterpretation often occurs during fixed-bias testing of close-spaced triodes such as the nuvistor. The opposite change, caused by an increase in work function of the grid wires (ψg) during life, has similarly been misinterpreted as emission slump instead of as instability of the grid work function.

Test conditions which separate the effects of grid work function and cathode work function are, therefore, desirable. Well-designed practical circuits have built-in regulating functions (e. g., feedback or degeneration) which compensate for potential changes, but can not restore cathode-emission failures.



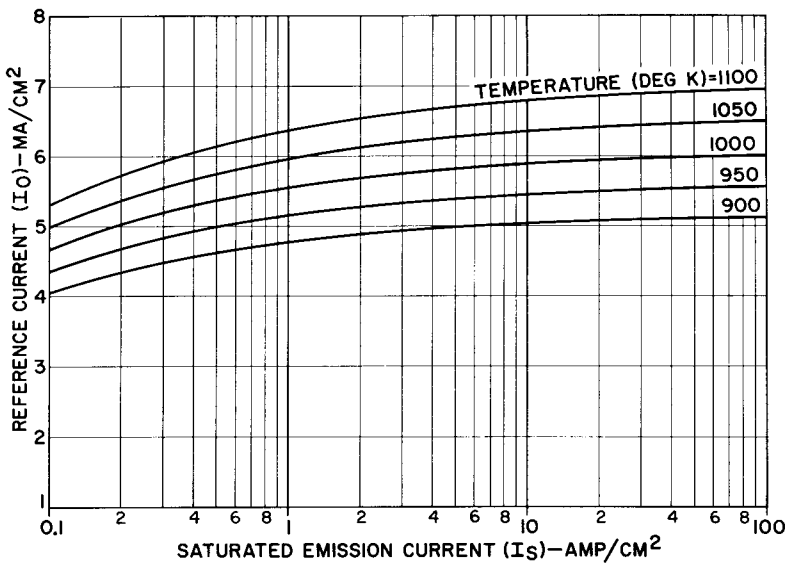


Figure 7. The Reference Current I_0 (Exponential point) as a Function of Emission Currents I_s and Temperature T for Two Fixed Cathode-to-Anode Distances d: (A) d = 2 mils; (B) d = 2.5 Mils

Table IV

I and g as f(T) for perfect diode; d = 0.002 inch, ψ_c = 1.6 volts, and V = (E_b - ψ_a + 2.216) + ΔV_o

	(E_b - ψ_a + 2.216) = 0.262						(E_b -ψ_a + 2.216) = 0.45				
T deg K	ΔV_o	V	I	g	per-cent I_{1100}	per-cent g_{1100}	V	I	g	per-cent I_{1100}	per-cent g_{1100}
1100	0	0.262	45	148	100.	100.	0.45	74	162	100	100
1050	-0.1	0.162	29	130	64.5	87.9	0.35	55	150	74.4	92.5
1000	-0.2	0.062	15	110	33.3	74.4	0.25	38	133	51.4	82.

TRIODES

It has been shown that the space current in a diode is determined in principle by the total charge Q = CV of the diode capacitance divided by the transit time (τ) during which the charge is moved through the distance d between cathode and anode. The problem of calculating the space current in electron tubes having a more complex electrode configuration (triodes, pentodes, and the like) can therefore be reduced to the diode case by evaluating an effective mean charge $\overline{Q} = \overline{CV}$ in the plane of the electrode nearest to the cathode. Calcu-

lation of a mean effective capacitance $\overline{C}$ and mean effective potential $\overline{V}$ in the plane of the first electrode (grid No. 1 in a triode) requires knowledge of the total electrostatic field in the space between the cathode and the first electrode. In the classic approach, this field is treated as a linear superimposition of the partial fields from all electrodes penetrating into this plane. Each field component is evaluated and averaged separately by calculation of penetration factors, $D_{n,1} = 1/\mu_{n,1}$, which specify the contribution of the potential field from electrode n to the total average potential $\overline{V}_1$ in the plane of the first electrode, so that a mean effective potential $\overline{V}_1$ can be obtained by a simple summation.

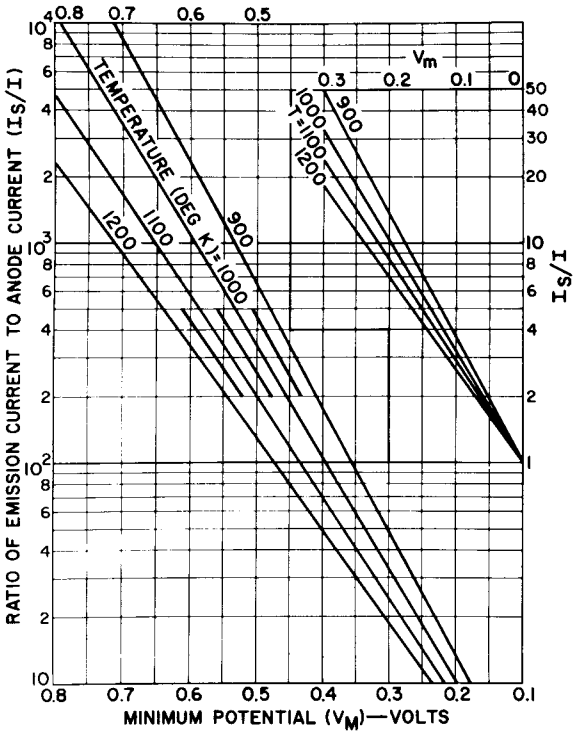


Figure 8. Minimum Potential $V_m = f(I_s/I)$

The n-1 basic potential field components in the three-dimensional space (x, y, z,) between the cathode and the first element of an electrode structure with n elements are the unit-potential fields $V_{n(1)} = f(x, y, z)$ of the electrodes. These fields can be calculated (or measured) by applying one potential unit (one volt) to the specified electrode (e. g., the grid in a triode) and zero potential to all other electrodes. The solution of the generalized case is rather difficult, so that classic solutions have been limited to electrode fields which are uniform in the z-dimension. The unit-potential fields can be described in this case by two-dimensional functions $V_{n(1)} = f(x, y)$, as shown in Fig. 12 for a triode. The penetration factor $D_{a,1}$ for the anode field can be visualized as the ratio of the number of electrostatic anode flux lines (broken lines in Fig. 12A) penetrating the grid plane and terminating on the cathode to the number of grid flux lines. The calculation of penetration factors ($D = 1/\mu$) is thus, in principle, a calculation of the ratios of element capacitances. The brief discussion below illustrates their use in deriving the classic expression for the equivalent grid potential of a triode, and points out a number of assumptions which limit the accuracy of the equivalent grid potential when it is applied to practical close-spaced triodes.

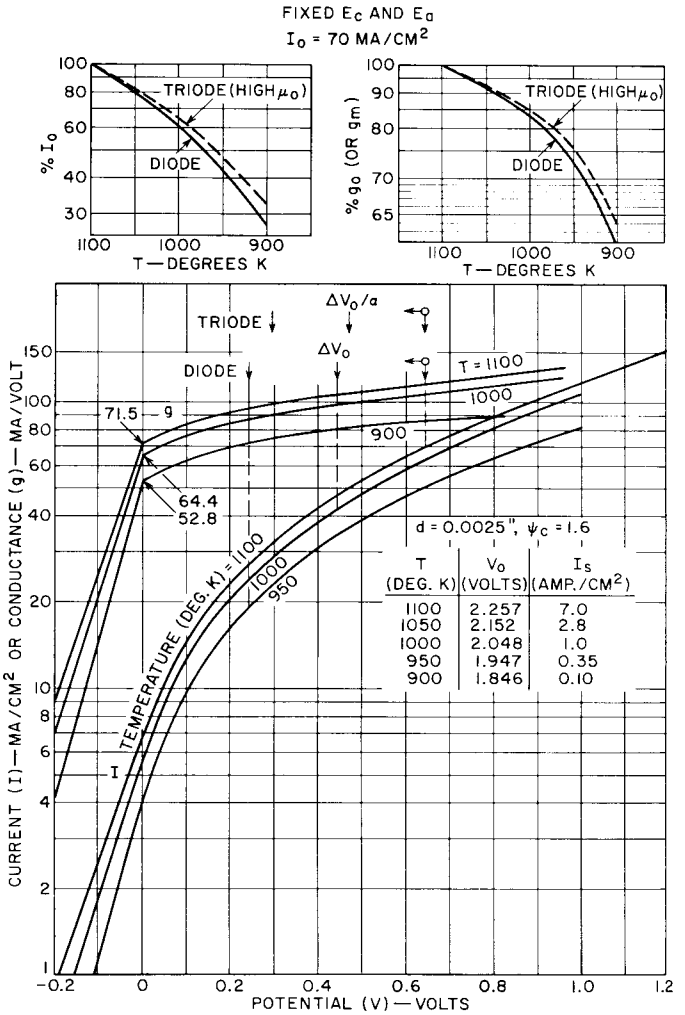


Figure 9. Diode Current I and Conductance g for d = 2.5 Milliinches, $\psi_c = 1.6$ as Functions of V and T. (Insets show change of current (I/I_{1100}) and conductance (g/g_{1100}) from initial values I_{1100}, g_{1100} obtained with fixed electrode potentials at T = 1100 K when T is reduced to 900 K.)

Electrostatic Capacitances and Potentials in "Perfect" Parallel-Plane Triodes (μ = Constant)

The capacitances between triode elements are shown in Fig. 13. The total capacitance C corresponding to the sum of all flux lines between the grid plane and the cathode (non-emitting, cold) is given by

$$C = C_g + C'_a \tag{20}$$

where $$C'_a = C_a(d_1 + d_2)/d_1 = C_a\left(1 + \frac{d_2}{d_1}\right) \tag{21}$$

The assumption that the capacitances C'_a and C_a are inversely proportional to the distances d_1 and $d_1 + d_2$ becomes inaccurate for the relatively thick grid wires and coarse pitch of practical grids, as will be shown

in the second part of this article. The penetration factor D is, by definition, a flux or capacitance ratio:

$$D = 1/\mu = C_a/C_g \tag{22a}$$

By means of Eq. (21):

$$C_a'/C_g = D(1 + \frac{d_2}{d_1}) = (1 + \frac{d_2}{d_1})/\mu \tag{22b}$$

The total charge Q due to applied potentials V_g and V_a without space charge (tube cold) is given by

$$\begin{aligned} Q = CV &= C_g V_g + C_a V_a \\ V &= \frac{C_g}{C} V_g + \frac{C_a}{C} V_a \end{aligned} \tag{23}$$

Substituting the expression for C from Eq. (20):

$$V = \frac{C_g}{C_g + C_a'} V_g + \frac{C_a}{C_g + C_a'} V_a$$

$$= \frac{V_g + (C_a/C_g) V_a}{1 + C_a'/C_g}$$

Then from Eq. (22):

$$V = \frac{V_g + V_a/\mu}{1 + \frac{1}{\mu}(1 + \frac{d_2}{d_1})} \tag{24}$$

Correction for the Space Charge by an Equivalent Diode Spacing (d_1')

To correct the electrostatic potential V for the effect of the space charge in the cathode-grid space in first approximation, the distance d_1 is decreased to an equivalent value d_1' which, in a parallel-plane diode, would give a constant potential gradient in the space-charge-free model equal to the actual potential gradient at the diode plate. Fig. 14 shows that straight lines having the slope of the actual space-potential function intersect the reference potential V = 0 at a shorter distance d' which is a function of V. The ratio $\rho = d_1/d_1'$ of the actual distance to the equivalent cathode distance can be approximated as follows:*

$$\begin{aligned} \rho = d_1/d_1' &= 3/2 \text{ for potentials } 1.5 < V > 0.1 \text{ volt} \\ \rho = d_1/d_1' &= 4/3 \text{ for potentials } 1.5 \quad V > 1.5 \text{ volts} \end{aligned} \tag{25}$$

Eq. (24) can then be replaced by the following expressions:

$$\begin{aligned} V^+ &= (V_g + \frac{V_a}{\mu}) \frac{1}{\alpha} \\ \alpha &= 1 + \frac{1}{\mu}(1 + \rho \frac{d_2}{d_1}) \end{aligned} \tag{26a}$$

The remaining potentials of the triode circuit (i. e., the work functions ψ_g and ψ_a of the grid wires and anode, the external applied voltages E_c and E_b, and the barrier potential V_0) are shown in Fig. 13. The grid and anode potentials are given by

*For further discussion see the second part of this article and Fig. 24.

$$\begin{aligned} V_g &= V_0 - \psi_g + E_c \\ V_a &= V_0 - \psi_a + E_b \end{aligned} \tag{26b}$$

The equivalent diode potential of the triode containing a perfect grid is, therefore,

$$V = \left\{(V_0 - \psi_g + E_c) + \left[(V_0 - \psi_a + E_b)/\mu\right]\right\} \frac{1}{\alpha} \tag{27}$$

$$\alpha = 1 + \frac{1}{\mu}(1 + \frac{\rho d_2}{d_1})$$

The total cathode current $I_k = I_a + I_g$ of the perfect triode can hence be calculated from V, as described for the perfect diode, and the transconductance $g_{m(k)}$ of the perfect triode in the cathode circuit (or without grid current in the anode circuit) is given by the diode conductance multiplied by the factor $1/\alpha$:

$$g_{m(k)} = g/\alpha \tag{28}$$

The relative changes $\Delta I/I$ and $\Delta g/g$ discussed for the diode as functions of T have, therefore, the same values for perfect triodes for constant current and constant grid potential V as they have for perfect diodes. The changes for fixed-voltage operation are reduced because ΔV_0 in Table IV is replaced by $\Delta V_0/\alpha$. (See also Fig. 9.)

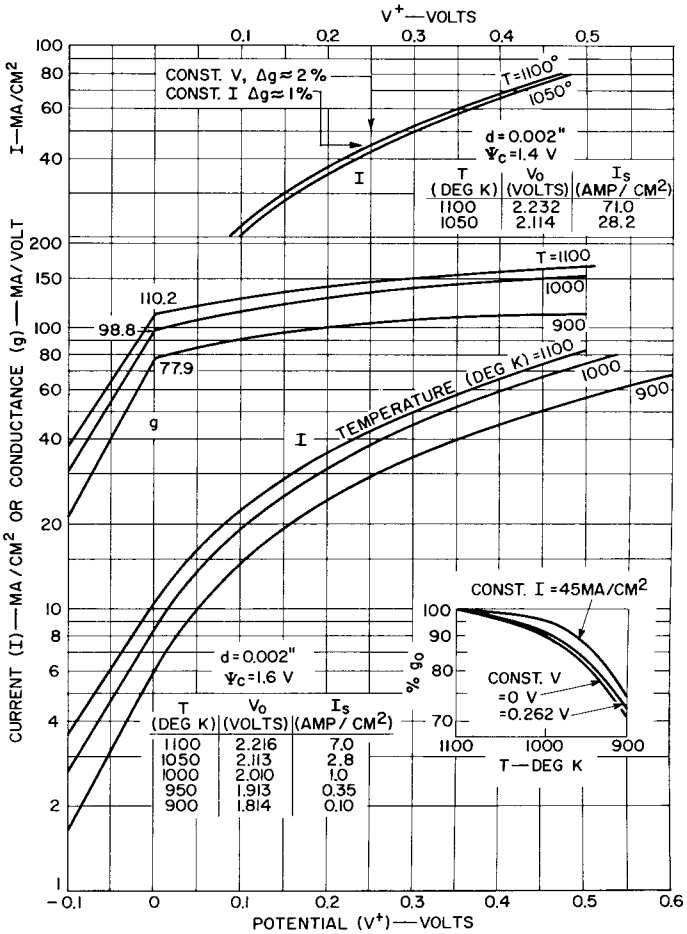


Figure 10. Diode Current I and Conductance g for d = 2 Milliinches as Functions of V and T. (Inset shows conductance change g/g_{1100} for a constant current I or constant potential V as a function of T.)

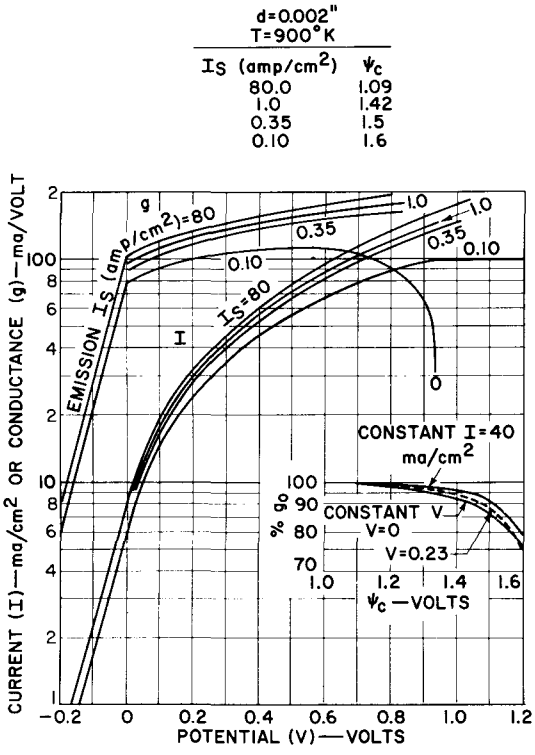


Figure 11. Diode Current and Conductance I, g = f(V) for Different Work Functions (ψ_c) and Emission Currents (I_S) for d = 2 Milliinches, T = 900 K (Inset shows conductance change $g/g_{1.1}$ for ψ_c = 1.1 to ψ_c = 1.6.)

Aberrations in Practical Triodes and Limitations of the Simple Theory

An equivalent diode potential [Eq. (27)] has been established for the triode by the addition of the mean potentials derived from the electrostatic potential fields of the grid wires and the anode (see Figs. 12 and 13) with a correction for an average space-charge effect in the grid-cathode space by the introduction of an equivalent cathode-grid distance $d' = d/\rho$. It is assumed that the grid provides a substantially constant electrostatic gradient so that the sum of the charges over the cathode $C_gV_g + C_aV_a$ [Eq. (23)] is sufficiently uniform and the current density at the cathode is reasonably constant. With practical grids, this condition can be approached only at certain ratios V_a/V_g. The field plots in Fig. 12 illustrate the basic electrostatic-field components obtained with a relatively good grid (practically speaking). These components are the unit anode field (Fig. 12A) obtained with one volt (100 per cent) applied to the anode and zero volts on grid and cathode, and the unit grid field (Fig. 12B) obtained with one volt on the grid and zero volts on anode and cathode. All other possible fields are obtainable from these unit fields by multiplying the space potentials of the unit fields by appropriate voltage factors and then adding the two potential components for each point in space. The field for V_a = 100 and V_g = -4.5, shown in Fig. 12C, was constructed in this manner.

Fig. 12A shows that even for V_g = 0 (i.e., with grid current) the field gradient is somewhat lower under the grid wire than between grid wires. The gradient variation increases considerably when a negative field is superimposed (see Fig. 12C). The gradient directly under the grid wire becomes negative, and causes a reduction of the active emitter area. This area contracts towards the center between grid wires as V_g is made more and more negative. The conditions for negligible grid current require a retarding field between grid wire and cathode without potential minimum (compare Fig. 14), and produce a potential minimum along the center line between grid wires, where the space charge "bulges" out from the cathode to form a potential "saddle," as estimated in Fig. 12D for a certain plate current.

It follows that for negative grid potentials ($V_g < 0$) the potential V computed from Eq. (27) is too low because the field distortion causes a decrease of the "active" emitter area by a factor $A/A_0 < 1$, reduces the "active" part of the capacitances C_g and C_a, and alters their ratio $\bar{\mu}$ by a factor $\bar{\mu}/\bar{\mu}_0 < 1$. The problem of computing the ratios A/A_0 and $\bar{\mu}/\bar{\mu}_0$ from the triode dimensions as functions of the independent variables V_g and V_a has not been solved analytically, although the distribution of electrostatic point $\bar{\mu}$ values (no space charge), i.e., the ratio $\mu = V_{g(1)}/V_{a(1)}$ has been computed for any point in the cathode-grid space. The normal electrostatic value μ_0 computed from grid dimensions and spacing is the ratio of average potential values, while the "point- $\dot{\mu}$" function has a maximum under a grid wire and a minimum under the center between grid wires. For parallel-plane structures this $\dot{\mu}$-function is different in different planes below the grid, but the average value $\bar{\mu}_0$ is constant for any plane in the grid-cathode space parallel to the grid plane. The "active" average μ value (i.e., the electrically measured μ value), is a particular average in a surface of the grid-cathode space in the presence of space charge, i.e., an average weighted according to the plate-current contribution. Its calculation is, therefore, not possible from electrostatic values without knowledge of the current-distribution function, as shown in the second part of this article.

It is, therefore, important to know whether there are operating potentials which result in a close approach to the theoretical values $A/A_0 = 1$ and $\bar{\mu}/\bar{\mu}_0 = 1$, which can serve as reference and check-values. The answer is "yes" for positive grid potentials near or equal to the potential in the cathode-anode space which would exist if the grid were removed. This condition is called "space-potential operation."

The Grid at Space Potential

The field plot in Fig. 15 shows that a practical triode having a grid with relatively large wires can be operated at a positive grid potential, which causes a minimum field disturbance in the grid-cathode space. The actual grid voltage is somewhat below space potential to permit bending of the plate field around the grid wire into the space between grid wires. Because of the uniformity of the parallel-plane field below the grid, the current intercepted by the grid is given by its shadow area; i.e., I_g/I_k = shadow area of grid/total area of

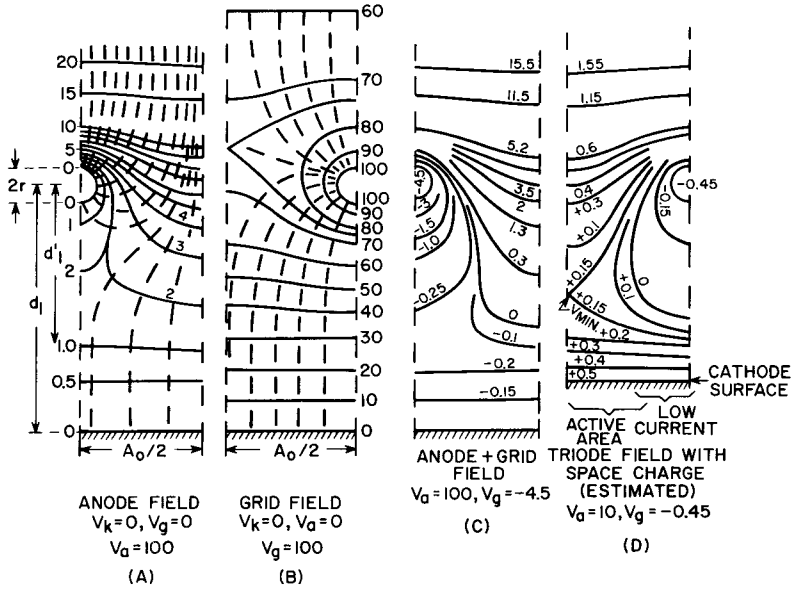


Figure 12. Unit Potential Fields and Composite Potential Field of a Triode: (A) unit anode-field $V_k = 0$, $V_g = 0$, $V_a = 1$ volt (100 per cent); (B) unit grid-field $V_k = 0$, $V_a = 0$, $V_g = 1$ volt (100 per cent); (C) composite electrostatic field for $V_a = 100$, $V_g = -4.5$; (D) composite field with space charge (estimated) for $V_a = 10$, $V_g = -0.45$

grid. This fact can be used to establish space-potential operation, i. e., to adjust the plate (or grid) voltage until this current ratio is obtained. *

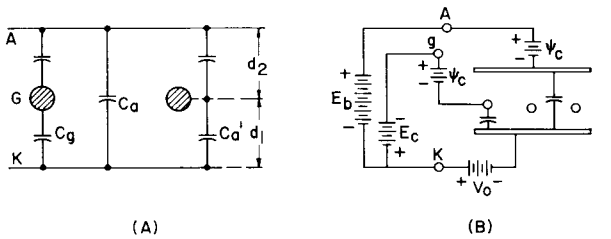


Figure 13. Triode Capacitances and Potentials

In Fig. 15, the equivalent-diode plane (V = 9.3) lies three-fifths of the grid-wire radius below the center line of the grid. The cathode current should thus coincide with the theoretical diode current computed for this diode spacing (slightly less than d) and the full cathode area A_o. Because of the relatively heavy grid current, the null indicator in bridge measurements of μ or g_m must be placed in the cathode circuit if it is to indicate the theoretically computed values. Actual measurements on pencil tubes and nuvistors have given close agreement with the values computed for the "perfect" triode when the remaining triode parameters were substantially uniform. An example of g_m-measurements on a nuvistor triode with 60-siderod, 0.85-mil-wire grid is shown in Fig. 16. The space-potential curve, labelled $I_g = I_k/3$, is very close to the curve for a perfect triode with d = 2.55 mils in the accelerating-potential region (V > 0), where the electrostatic value $\bar{\mu}_o = 35$ is also obtained.

*An analytic method is given in the second part of this article.

The remaining g_m curves are reasonably parallel to the theoretical curve; the parameter is a constant μ value (measured). The broken lines indicate grid-bias values. The bias for a grid current of 0.5 microampere is approximately -0.4 volt at I_b = 10 milliamperes and 0.5 volt at I_b = 1 milliampere. (The curves are not continued into the grid-current region toward space-potential operation because the present μ bridge does not have a null indicator in the cathode circuit.)

In the retarding field V < 0 (below the knee point), true space-potential operation can only be approached because a positive gradient at the anode side of the grid is required to keep electrons which pass between the grid wires from returning to the grid. The best condition, shown in Fig. 17, is obtained for equal and opposite gradients. It produces a saddle field and a considerable nonuniformity even near very small grid wires, so that the sharp knee of the g_m function cannot be obtained with practical grids in close-spaced tubes. When V is made more negative, the minimum disappears and the measured curve approaches the theoretical curve asymptotically. It is, therefore, not possible to use the slope or position of the measured retarding-field functions of g_m or I_b as temperature indicators except at very low currents.

The Grid Work Function ψ_g

Calculation of effective grid potentials requires knowledge of the grid and anode work functions. Exact knowledge of the anode work function is not critical because ψ_a is divided by μ, but the grid work function ψ_g must be known accurately to at least 0.02 volt. The grid work function ψ_g can be computed from Eq. (27):

$$-\psi_g = \alpha V - (V_o + E_c) - \left[(V_o - \psi_a + E_b)/\bar{\mu}_o\right] \quad (29)$$

where

$$\alpha = 1 + \frac{1}{\bar{\mu}_o}\left(1 + \rho\frac{d_2}{d_1}\right)$$

According to the preceding comments, this equation will furnish precise values only for perfect triode conditions, i.e., for the theoretical maximum value $\bar{\mu}_o$ and an active cathode area equal to the total actual area A_o of the cathode. The measurement should therefore be made at or close to space-potential conditions for the grid, where $\bar{\mu} = \bar{\mu}_o$ and $A/A_o = 1$.

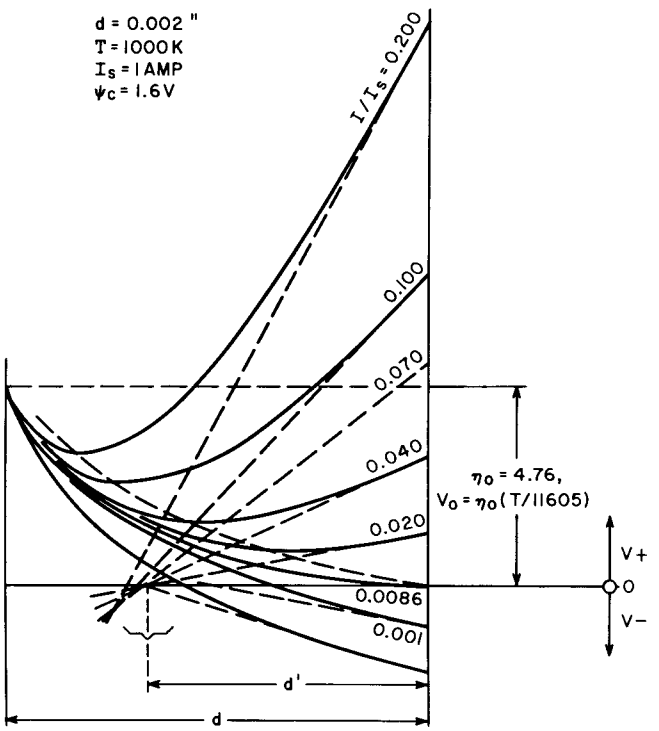


Figure 14. Space-Potential Functions and Equivalent Distances d' of a Diode as Functions of Effective Potential V

Cathode-Temperature Variations ± ΔT

The thermal circuit of oxide-coated cathodes consists essentially of three shunt paths from the cathode to the heat sink which receive a large fraction K of the total heater power P_h:

$$KP_h = P_e + (P_r + P_c) \qquad (30)$$

The remaining fraction (1-K) of the power is the direct heater end-loss to the heat sink.

The power loss P_e is given by:

$$P_e = I(\psi_c + V_m) + 2V_T \text{ watts} \qquad (4)$$

This loss represents the thermal energy expended by the electrons in the current I in acquiring the velocity for emission and in overcoming the potential barrier $(\psi_c + V_m)$. The bracketed term in Eq. (30) is given by:

$$(P_r + P_c) = c(T_c^4 - T_2^4) + (T_c - T_s)/R \qquad (31)$$

This term consists of the total power P_r radiated from the cathode sleeve at a temperature T_c to surrounding electrodes having a temperature T_2 (c is a constant), and a conduction loss P_c in the thermal resistance R of the cathode-support member connected to a heat sink of temperature T_s.

The removal of electrons by a current I from the cathode causes a cooling effect which decreases the cathode temperature T_c by the same amount as a heater-power reduction $\Delta P_h = P_e/K$. Given the factor K and the function $T_c = f(P_h)$ (shown in Fig. 18 for a nuvistor triode), the temperature change - $\Delta T_c = f(I)$ is easily computed for a given work function and cathode temperature. For K = 0.88, T_o = 1100 K, ψ_c = 1.6 volt, and a current I = 22 milliamperes, for example, the power loss due to electron removal is $P_e/0.88 \simeq 0.05$ watt, which reduces the cathode temperature by ΔT = -22 K, as shown in Fig. 18. This temperature change causes a potential change ΔV_o = -0.044 volt which, for fixed-bias operation, results in a current reduction $I_D \simeq \Delta V_o \times g \simeq$ -7 per cent for a perfect diode (see Fig. 11). For triodes, the potential change is $\Delta V_o/\alpha$, which results in $\Delta I_{T_r} \simeq$ -6 per cent = -1.3 milliamperes for the example discussed.

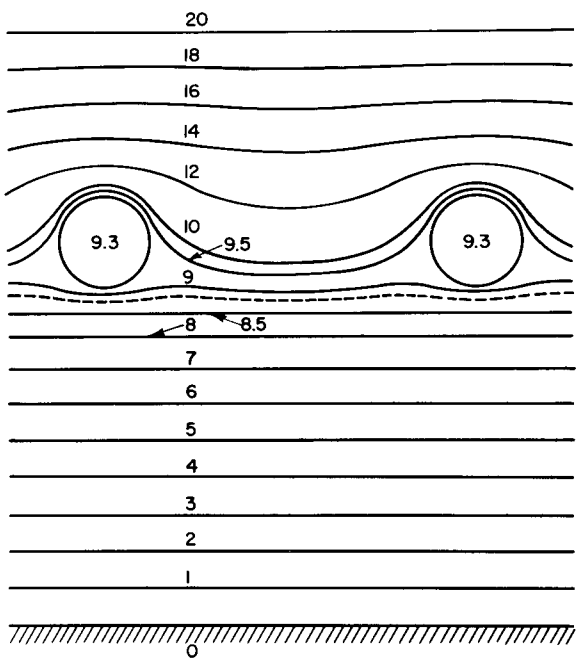


Figure 15. Potential Field of Triode with Grid at Space Potential (V positive)

The cooling effect - $\Delta T = f(\bar{I})$ and the relative current change $\bar{I}/\bar{I}_o = f(\Delta T)$ have been measured on a nuvistor having a thermocouple welded to the cathode. The results are shown in Fig. 18. A current I = 22 milliamperes causes ΔT = -25 K, and this change causes $\bar{I}$ to decrease nearly 7 per cent, which is slightly more than that computed. The measurements show that there is a difference between a direct current and a modulated current having the same average value. Because the

modulated current has a peak current considerably higher than its average value, it causes a larger change $-\Delta T$, and ΔT in turn causes a larger change $-\Delta \bar{I}$ than a direct current. This difference indicates nonuniform cooling and perhaps a complex thermal time constant having a rapid decay component besides the slow main component, which is of the order of several seconds. The generally nonuniform spacing and current density of practical tubes certainly contributes to a nonuniform cooling effect.

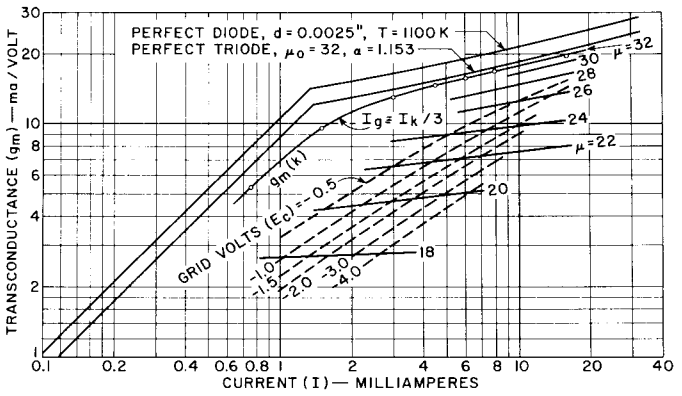


Figure 16. Theoretical and Measured Transconductances (g_{mk}) of a Triode ($d_1 = 0.00255$ inch, $d_2 = 0.00665$ inch, $\bar{\mu}_o = 32$, $A = 0.185$ cm^2)

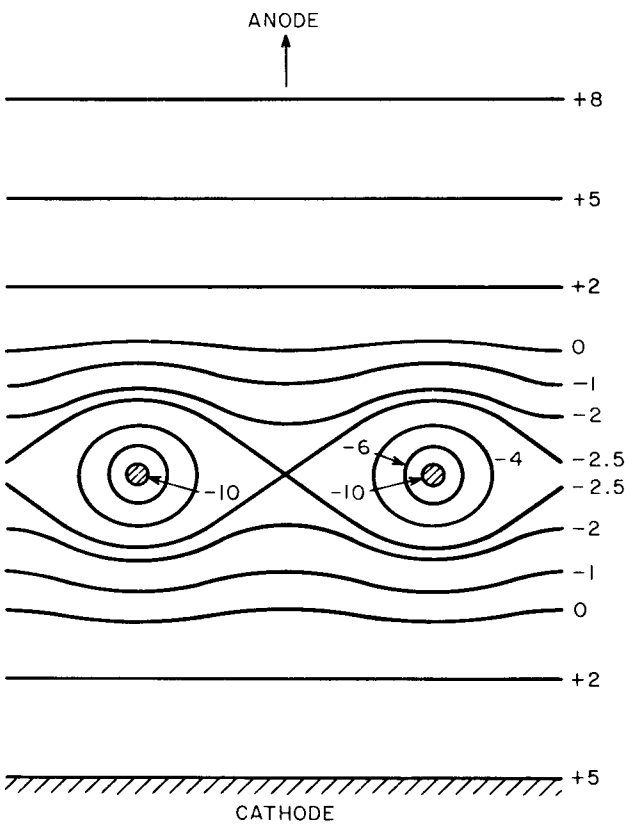


Figure 17. Potential Field of Triode with Grid at Space Potential (V negative)

It is evident that the initial current drift caused by emission cooling does not indicate a processing fault. It is more pronounced in close-spaced tubes and will increase for cathodes having a reduced power loss. The appreciable magnitudes of $-\Delta T$ and $-\Delta \bar{I}$ can explain many discrepancies observed between static and dynamic measurements (curve tracer) of tube characteristics.

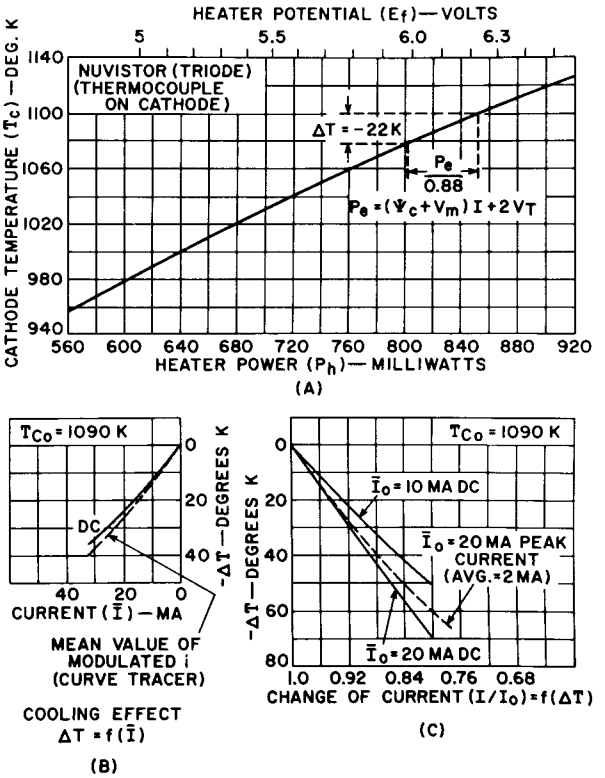


Figure 18. Cathode Temperature (T_c) and Cooling Effects ($-\Delta T$) Caused by Electron Emission: (A) cathode temperature of a triode as a function of heater power ($I = 0$); (B) cooling effect $-\Delta T$ as function of average current; (C) current change $\bar{I}/\bar{I}_o$ as function of temperature change

A positive temperature change $+\Delta T$ can be observed when the temperature T_2 of surrounding electrodes increases substantially due to electrical power dissipation. Eq. (31) shows that the radiated power P_r decreases when T_2 rises; the actual relation between T_c and T_2 is considerably more complicated in a triode, and the grid shielding action permits a much higher plate temperature (T_2) than indicated by Eq. (31) before a noticeable reduction of P_r and a consequent temperature increase $+\Delta T_c$ of the cathode is observed. Measurements on a nuvistor triode have shown substantially no change of T_c when the current is held constant and the plate dissipation is increased from a very small value to 1 watt by adjustment of E_c and E_b.

Eccentric Electrodes

Uniform Eccentricity of Grid and Cathode. The spacing variation $\Delta d_1 = f(\phi)$ is sinusoidal. The current distribution along the cathode periphery can be approximated by $i/i_o = [1/(1-\Delta d_1)]^2$, as shown in Fig. 19.

The table in the figure shows the average current ratio $\bar{I}/\bar{I}_0$, as well as the maximum value $I_{max}/\bar{I}_0$, for various eccentricities $\pm \Delta d_{1_{max}}/d_1 = m_1$.

An increase of current density by a factor $I_{max}/\bar{I}_0 = 2$, for example, permits an eccentricity $m_1 = 0.3$. For the present nuvistor with $d \simeq 2.5$ mils, this eccentricity has the value $\pm \Delta d_1 = 0.75$ mils. The average current increases by $\bar{I}/\bar{I}_0 = 1.145$, or 14.5 per cent.

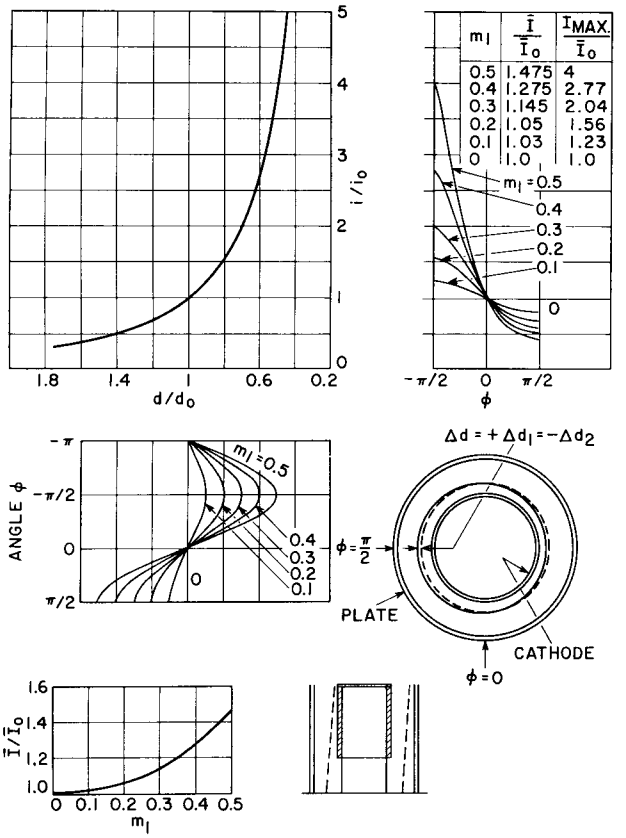


Figure 19. Graphic Evaluation of Current Change (i/i_0) Caused by a Sinusoidal Variation of Spacing (d/d_0) Due to Eccentric Electrodes. (Inset, left bottom: Increase of average current ($\bar{I}/\bar{I}_0$) as function of cathode grid eccentricity m_1.)

A good limit for eccentricity would be $m_1 = 0.2$; this value would result in $\pm \Delta d = 0.5$ mils, and would not permit the current density to exceed 1.56 times the normal value. The average plate-current tolerance would then be $\bar{I}/\bar{I}_0 = 1.05$, or zero to plus five per cent from bogie.

Eccentricity Caused by Grid Ellipticity. Eccentricity caused by grid ellipticity can occur when the grid spacer in a brazing jig is elliptical or has too small a diameter. In this case a double-frequency sinusoidal distortion occurs around the cathode periphery; this distortion produces the same result as uniform eccentricity.

Grid-Plate Eccentricity Δd_2. Variations Δd_2 can be caused by grid eccentricity or by an independent anode eccentricity. A change $\Delta d_2/d_2 = m_2$ causes an equal change $\Delta \mu/\mu$. Grid eccentricity simultaneously causes an eccentricity m_2 of opposite sign $m_2 = -m_1 (d_1/d_2)$. For a nuvistor having an 0.080-inch anode diameter, $d_1/d_2 = 0.38$, and a uniform grid eccentricity $m_1 = 0.2$ results in $m_2 = -0.076$ and $\Delta \mu/\mu = \mp 0.076$. The change in current i/i_0 caused by the sinusoidal variation of μ around the tube axis can be obtained from the static characteristic because $\Delta E_b/E_b = \Delta \mu/\mu = \mp 0.076$. At $E_b = 36$ volts and $I_b = 9$ milliamperes, the plate-voltage change in the nuvistor triode is $\Delta E_b = \mp 2.74$.* At constant E_c this small variation causes a substantially undistorted plate-current variation of sinusoidal form within the limits $\Delta_2 i \simeq \mp 1.3$ milliamperes or $\Delta_2 i/\bar{I}_0 = \mp 0.144$. Because of the opposing sign with respect to the change $\Delta i/\bar{I}_0$ caused by the grid eccentricity in the grid-cathode space, the combined change is obtained by subtracting the sinusoidal change $\Delta_2 i \sin \Phi$ from the current waveform labeled $m_1 = 0.2$ in Fig. 19. The resultant change in average current is $\bar{I}/\bar{I}_0 = +1.07$. The total change in average current is two per cent larger than that computed for the effect of Δd_1 alone, but the maximum current $I_{max}/\bar{I}_0$ is reduced from 1.56 to 1.42.

The effect of Δd_2 on the currents is smaller for nuvistors with larger plate spacing (higher μ).

An independent eccentricity of the anode can add to or subtract from the current changes. An increase of plate-current limits is not permissible because of the danger of excessive cathode loading.

Nonuniform Eccentricity (Tilting) of Cylindrical Electrodes. If a nuvistor mount is distorted during removal from a concentric jig, the grid cylinder may be eccentric to the cathode and anode at only one end (see sketch in Fig. 19). If $m_1 = 0.3$ at this end and zero at the opposite end, the maximum current density is increased by a factor $\bar{I}_{max}/\bar{I}_0 = 2$. The mean current change is obtained by averaging the values $\bar{I}/\bar{I}_0$ in the axial direction, i.e., the function $\bar{I}/\bar{I}_0 = f(m_1)$ shown in Fig. 19 up to $m_1 = 0.3$. This total average is $\bar{I}/\bar{I}_0 = 1.05$, i.e. a plate-current tolerance plus five per cent, minus zero per cent from bogie. A plus ten per cent tolerance would permit $m_1 = 0.45$ and $\Delta d_1 = 1.12$ mils as maximum values and an increase of 3.4 in maximum current density. Therefore, to obtain good life expectancy the maximum eccentricity in a nuvistor should not exceed $\pm \Delta d_{1_{max}} = 0.5$ mils for uniform eccentricity along the cathode or $\pm \Delta d_{1_{max}} = 0.75$ mils for a tapered eccentricity. The corresponding permissible current change from bogie is plus five per cent, minus zero per cent when the test condition assures a constant effective grid potential. These permissible eccentricities are certainly obtainable values.

Calculation of Vibration-Induced Modulation in Cylindrical Structures

A sinusoidal mechanical vibration of the cantilever-supported grid in a nuvistor causes an electrical sine-

*The reader may compare these increments with those taken from the machine-computed characteristic of a similar nuvistor shown in Fig. 40.

wave output signal of double frequency when the spacings d_1 and d_2 are uniform. Eccentricity of the mount causes signal components of fundamental frequency. It follows from the transfer characteristic $i = f(d_1)$, shown in Fig. 19, that vibration of a concentric or eccentric grid causes a sinusoidal variation of m_1, which in turn causes a variation $+\Delta\bar{I}$ of the current average $\bar{I}$, thus producing an electrical modulation of $\bar{I}$.

The effect of a vibration of fixed amplitude is illustrated in Fig. 20 for several eccentricities. Cases (B) and (C) represent uniform eccentricities along the axis at zero vibration amplitude, i. e., with grid and cathode axes parallel and displaced. If it is assumed for simplicity that the cathode coating lies between the halfway point of the total cylinder length and the top of the cylinder, the top-point amplitude is approximately double that of the half-length points. This assumption is not strictly correct, but can serve to determine the type of output signal obtained. The top amplitude is assumed to be a 20-per-cent spacing modulation in all cases illustrated, i. e., $\Delta m = 0.2$. Since the bottom points of the cathode receive half this modulation, i. e., $\Delta m = 0.1$, the total current from all cathode points is obtained as a double integral, i. e., the average value of all average values $\bar{I}/\bar{I}_0$ along the cathode axis from bottom to top. The range of m values is indicated in (A) to (C) for the positive, zero, and negative amplitudes. The integrated values $\bar{I}/\bar{I}_0$ (from Fig. 19) produce the waveforms shown in Fig. 20:

Fig. 20A shows the waveform for a perfect coaxial grid-cathode assembly, for which the electrical output is a double-frequency modulation. Fig. 20B shows the result of a uniform eccentricity (m = 0.2) along the cathode which produces a large single-frequency component in the output. Fig. 20C shows the result of a smaller uniform eccentricity (m = 0.1) causing a much smaller single-frequency component approaching the amplitude of the double-frequency component. Fig. 20D shows a condition obtained by adding a tilt to the case shown in Fig. 20C so that the half-length eccentricity becomes zero. The output waveform contains a smaller fundamental component.

More accurate calculations of this type should establish good numerical correlation between observed waveforms and amplitudes and the actual eccentricities of nuvistor structures, provided the Q values of the mechanical grid resonance are reasonably constant.

A REFINED METHOD FOR CALCULATING THE SPACE CURRENT DISTRIBUTION AND CURRENT-VOLTAGE CHARACTERISTICS OF ELECTRON TUBES

A NEW APPROACH TO SPACE-CURRENT CALCULATION IN ELECTRON TUBES

The classic solution for the space current in a triode makes use of penetration factors ($D = 1/\mu$) derived from the electrostatic fields of the electrodes to establish a mean equivalent diode potential $V_e = (V_g + D_a V_a)/(1 + D_a + 4/3\ D_k)$ in order that the space current may be computed from the Langmuir relations.[2, 3, 4] The effect of space charge in the cathode-grid space is approximated by a 4/3 increase of the reverse penetration factor (D_k) of the cathode field through the grid. This solution is valid when the composite field and space current in the cathode-grid space are substantially uniform. The assumptions, however, result in intolerable errors in modern close-spaced electron tubes, particularly at negative grid potentials.

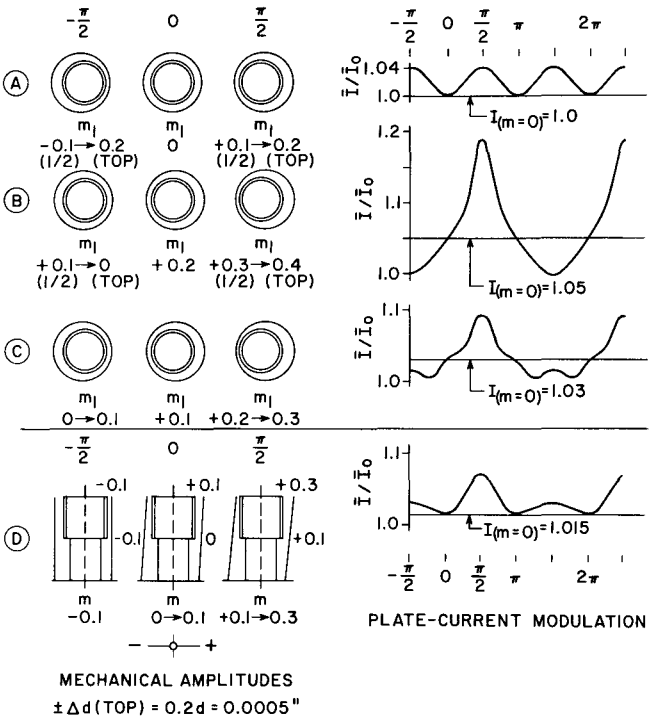


Figure 20. Vibration-Induced Current Modulation $\bar{I}/\bar{I}_0$ in a Nuvistor Triode; Relative Vibration Amplitude $m_{1max} = \pm 0.1$ at half length of grid, $= \pm 0.2$ at top of grid: (A) coaxial structure; (B) uniform eccentricity (m_1 zero = 0.2); (C) uniform eccentricity (m_1 zero = 0.1); (D) tilted eccentricity

It is fairly obvious that an accurate solution for the space current must be based on the detailed potential and current distributions in the grid-cathode space, taking into account the space charge and initial velocity of the electrons.

A rigorous analytic treatment of the electrode flow in non-uniform fields altered by space charge is exceedingly complex and will not be attempted because it requires a prohibitively large number of successive approximations to obtain a simultaneous solution for current and charge distributions, space potentials, and electron trajectories with appropriate initial velocities in the grid-cathode space. A study of the problem indicated that the tracing of electron trajectories in the space beyond 1/2 of the cathode-grid distance is of secondary importance in accelerating fields and may be dispensed with in a solution for the total current drawn from the cathode. It is not necessary to know where the electrons pass through the grid plane, provided the current distribution can be determined at the half-way distance in the grid-cathode space. The problem is thus reduced to a calculation of the effective field

strength (including space charge) in sections (y) taken normal to the cathode with the assumption of an independent parallel current flow in the second half of these sections. For a first approximation this assumption can be extended to the full cathode-grid distance. It will be shown subsequently that this simplification leads to errors only at highly retarding field conditions, i. e., at small currents near current cutoff.

The Electrostatic Field in the Cathode-Grid Space

The first step required in an accurate solution for the space current is the calculation of the electrostatic potential field in the grid-cathode space. The extreme complexity of a general solution is avoided by limiting the calculation of potential fields to parallel-plane triodes having a grid of uniformly spaced parallel wires; cylindrical triodes or structures having non-uniform spacings can be converted by transformation or segmentation into one or a number of equivalent parallel-plane triode sections.

The electrostatic potential field in the grid-cathode space of a parallel-plane triode is calculated from its geometric dimensions in normalized form using the grid-wire radius R as a unit. The electrode structure can hence be specified by the ratio of three dimensions X, Y, and C, where

$X = d_1/R$ = normalized cathode to grid-center distance

$2Y = P/R$ = normalized grid-wire pitch

$C = d_2/R$ = normalized grid-center to anode distance

The field calculation is made in two parts:

1) the unit potential grid field $v_{g(1)} = f(x, y)$

2) the unit potential anode field $v_{a(1)} = f(x, y)$ illustrated in Fig. 21.

These fields are computed for a matrix of 204 points by applying one volt to the grid or anode, respectively, with the remaining electrodes at zero potential. The total potential $v_{(x, y)}$ of a matrix point for any given combination of electrode potentials V_g, V_a is the sum

$$v_{(x, y)} = v_{g(1)} V_g + V_{a(1)} V_a = f(x, y) \qquad (32)$$

A typical set of unit potentials is given in Table V.

Amplification Factors (μ)

Although the electrostatic amplification factor $\bar{\mu}_o$ is not required for the evaluation of equivalent potentials, it is very useful as a normalizing factor for the anode voltage and is readily calculated from the unit potential functions.

The amplification factor $\dot{\mu}$ at a point in the grid-anode space is by definition the potential ratio

$$\dot{\mu} = v_{g(1)}/v_{a(1)} \qquad (33)$$

The mean electrostatic value $\bar{\mu}_o$ in the grid-cathode space (which is the object of normal μ calculations) is by definition the ratio of the mean unit potentials $\bar{v}_{g(1)}$ and $\bar{v}_{a(1)}$ in any plane $x < x_c$ in the grid-cathode space, where x_c is the plane just contacting the grid wires. The value $\bar{\mu}_o$ can therefore be calculated accurately from the potential data (Table V) with

$$\bar{\mu}_o = \bar{v}_{g(1)} / v_{a(1)}, \quad x < x_c \qquad (34)$$

As an example, it may be assumed that $x_c = 12$. For a given set of dimensions the electrostatic (mean) value $\bar{\mu}_o$ is independent of the distance x from the cathode within the restriction $x < x_c$, although the variation of point $\dot{\mu}$ values becomes larger with increasing x values. The value $\bar{\mu}_o$, however, is not independent of the model dimension X as generally assumed in $\bar{\mu}$ calculations.*

Fig. 22 shows the average potentials $\bar{v}_{g(1)}$ and $\bar{v}_{a(1)}$ (from Table V) as functions of x. The ratios $\bar{v}_{g(1)}/\bar{v}_{a(1)} = \bar{\mu}_o$ are constant for any value $x < 12$. For $x > 12$, some of the electrostatic flux lines in the space between grid wires do not penetrate the grid but terminate on the grid wires. These lines do not, therefore, contribute to the total field strength at the cathode.

Effective Potentials and Space-Current Distribution

The total field $v_{(x, y)} = f(x, y)$ is divided into 16 sections $v_{(y)} = f(x)$ which are treated as independent parallel triodes. The generally non-linear electrostatic fields $v_{(y)} = f(x)$ of these triodes are replaced by linear functions representing "equivalent" parallel-plane diodes so that the space current $i_{(y)}$ can be computed from the Langmuir relations. The total current is the sum of all section currents.

The evaluation of equivalent diode potentials for calculation of the section currents (i_y) requires several steps. It is well known that the electrostatic potential functions of a diode or triode are depressed by the negative space charge injected by an electron current, as shown in Fig. 23. The resultant gradient changes manifest themselves as an increase in the effective grid-cathode capacitance which can be duplicated in a space-charge-free model by a reduction of the normal grid-cathode distance d to an equivalent shorter distance d'. The distance d' for equal capacitance can be computed for a theoretically perfect parallel-plane diode or triode as follows: The potential gradient dv/dx at the anode of the actual potential function with space charge is matched by the constant gradient line of a space-charge-free diode (broken lines). The intersection of the constant gradient lines for two slightly different anode potentials occurs at the distance d'.[2] Because all intersections occur close to the potential level of the exponential point for which the anode gradient is equal to zero, d' can be determined with good accuracy from the intersection of one gradient line with the fixed potential level of the exponential point. The potential

*The majority of classic $\bar{\mu}$ calculations assume an infinite cathode-grid distance, and give an asymptotic value.

Table V

Electrostatic Space Potentials (Unit Potentials) for X/Y/C = 4/4/28

Plate Field $V_{a(1)} \times 10^2 = f(x, y,), \quad \bar{\mu}_o = 56.3$									
y =	0	2	4	6	8	10	12	14	16
x = 2	0.300	0.294	0.276	0.248	0.214	0.180	0.149	0.128	0.121
4	0.613	0.600	0.561	0.503	0.430	0.355	0.288	0.242	0.226
6	0.952	0.931	0.868	0.771	0.650	0.521	0.406	0.326	0.298
8	1.335	1.30	1.21	1.06	0.877	0.674	0.487	0.357	0.310
10	1.780	1.74	1.61	1.40	1.12	0.809	0.511	0.300	0.227
12	2.310	2.26	2.09	1.81	1.41	0.934	0.441	0.095	0.0
14	2.960	2.89	2.68	2.32	1.80	1.09	0.229	0.0	0.0
16	3.74	3.66	3.43	3.01	2.37	1.43	0.0	0.0	0.0
Grid Field $V_{g(1)} = f(x, y)$									
y =	0	2	4	6	8	10	12	14	16
x = 2	0.106	0.106	0.109	0.113	0.119	0.125	0.130	0.134	0.135
4	0.209	0.211	0.217	0.226	0.238	0.250	0.262	0.271	0.274
6	0.309	0.311	0.321	0.335	0.354	0.376	0.397	0.413	0.419
8	0.403	0.407	0.419	0.440	0.468	0.501	0.537	0.566	0.577
10	0.489	0.494	0.510	0.536	0.575	0.625	0.683	0.737	0.761
12	0.564	0.570	0.589	0.622	0.670	0.739	0.831	0.939	1.0
14	0.628	0.635	0.655	0.691	0.746	0.828	0.954	1.0	1.0
16	0.679	0.685	0.705	0.741	0.796	0.877	1.0	1.0	1.0

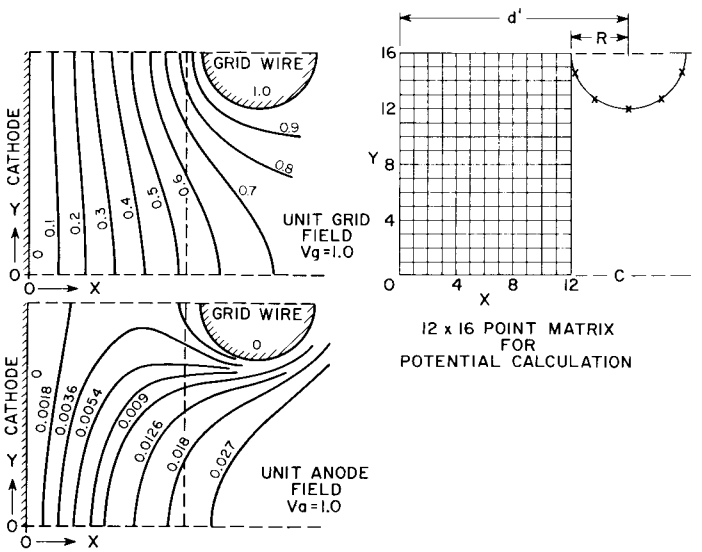


Figure 21. Unit-Potential Fields in the Cathode-Grid Space of a Triode of Table V

of the exponential point has, therefore been selected as the reference potential for all calculations:

$$V = 0 \text{ for } dv/dx = 0 \qquad (35)$$

Therefore, the cathode of the equivalent space-charge-free diode model has the convenient potential V = 0. Its spacing d' is a function of current (see Fig. 23) which can be derived from the Langmuir relations in normalized units:

$$\rho = d/d' = \xi_{(o)} (I/I_o) \gamma / \eta_v \qquad (36)$$

where $\xi_{(o)}$ = normalized distance d for $\eta_a = 0$

I_o = current at the exponential point

$\gamma = d\eta_a/d\ \xi_a$ = potential gradient* at the anode

$\eta_v = 11.605V$ = normalized diode potential (measured from the exponential point)

The reference ratio ρ_r is plotted in Fig. 24 as a function of the equivalent diode potential V_r for T_k = 1000 K,

*The negative gradients (γ) for retarding-field conditions can be found in Ref. 5.

$I_S/I_O = 1000$, $\xi_O = 2.509$. The scale factors for other temperatures and emission ratios are given by

$$\rho = \rho_r \, \xi_O/2.509 \text{ and } V = V_r \, T/1000 \quad (36a)$$

The space-charge correction of elemental sections in the actual triode requires that the cathode-grid distances d' of the electrostatic model used for calculations satisfy Eq. (36). The simultaneous solution for d' and V can be obtained with adequate accuracy by a linear interpolation from the equivalent potentials $V_{d'}$ calculated from two electrostatic models having the distance ratios $\rho_1 = 1.5$ and $\rho_2 = 1.8$, as indicated by the broken lines in Fig. 24.

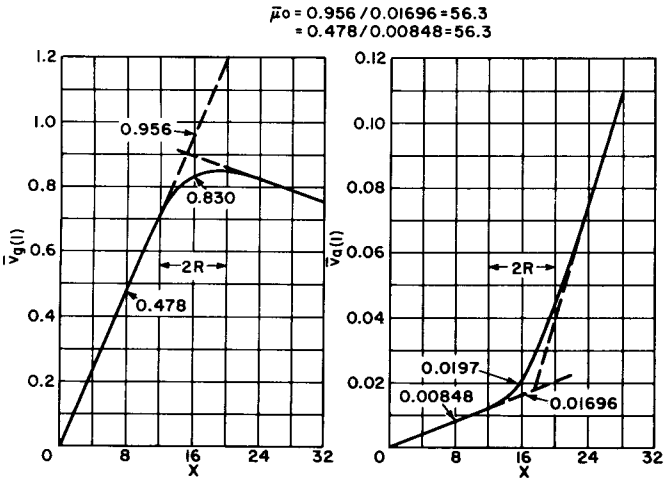


Figure 22. Mean Potentials $\bar{v}_{g(1)}$ and $\bar{v}_{a(1)}$ (From Table V) as Functions of x

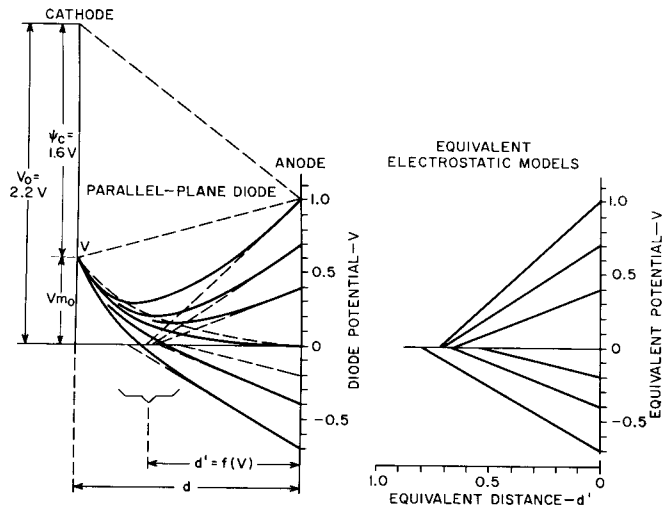


Figure 23. Space Potential Functions v = f(x) in a Parallel-Plane Diode (left) and Linear Potential Functions of an Equivalent Parallel-Plane Capacitance (right)

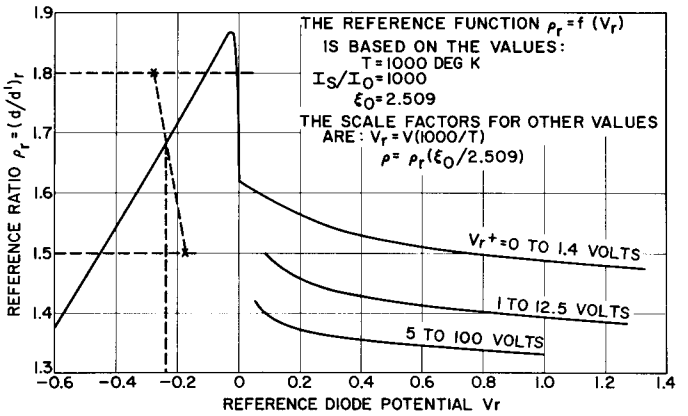

THE REFERENCE FUNCTION ρr = f (Vr)
IS BASED ON THE VALUES:
T = 1000 DEG K
Is/Io = 1000
ξo = 2.509
THE SCALE FACTORS FOR OTHER VALUES
ARE: Vr = V(1000/T)
ρ = ρr(ξo/2.509)
Vr+ = 0 TO 1.4 VOLTS
1 TO 12.5 VOLTS
5 TO 100 VOLTS
REFERENCE RATIO ρr = (d/d')r
REFERENCE DIODE POTENTIAL Vr

Figure 24. Equivalent Distance Ratio (ρ_r) as a Function of Effective Diode Potential (V_r)

To apply this solution to practical triodes, it is necessary to convert their nonlinear potential functions $V_{(y)} = f(x)$ (see Fig. 25) to constant gradient functions representing equivalent parallel-plane diodes. This conversion poses two problems: (1) where to locate the anode of the diode, and (2) how to replace the nonlinear function by an "equivalent" straight line. Basic considerations indicate that the anode of an equivalent diode should be located at the first "break point" of the electrostatic potential function, which for positive gradients occurs in the region between grid wires. (Compare Fig. 22). There are, however, potential functions exhibiting an electrostatic minimum (y = 8 in Fig. 25), which should be considered as a location for the anode of the equivalent diode. An extended numerical study indicated that a satisfactory equivalent can be obtained with a fixed anode location of the equivalent diode in the plane through the grid-wire centers. For this condition, the function $V_{(y)} = f(x)$ must be limited to a range between x = 0 and x = 12 (Fig. 21), which is the plane contacting the grid wires on the cathode side. The equivalent diode potential $V_{(y)}$ is extrapolated by placing a straight line through the origin and the average point $\bar{v}$ of the truncated function $V_{(y)} = f(x)$, as illustrated in Fig. 25. The correct combination of model distance d' and potential $V_{(y)}$ satisfying Eq. (36) is obtained by computing $V_{(y)}$ for two models having the distance ratios $\rho_1 = 1.5$ and $\rho_2 = 1.8$ and using the linear interpolation indicated in Fig. 24.

The diode currents and distribution i = f(y) can now be computed from the Langmuir functions from the potentials $V_{(y)} = f(y)$, as shown in Fig. 26 for the anode potential $V_a = \bar{\mu}_o$ and four grid-wire potentials for a 6CW4 nuvistor triode. The computed mean current densities and total currents for the particular cathode area (0.185 square centimeter) are indicated.

The equivalent electrostatic potential fields used in the calculation are shown in Fig. 27 for two operating conditions. The potential field in the actual cathode-grid space containing the space charge is obtained in first approximation by replacing the equivalent constant gradient functions of Fig. 27 by the corresponding space-charge functions (Fig. 25), as shown in Fig. 28. For

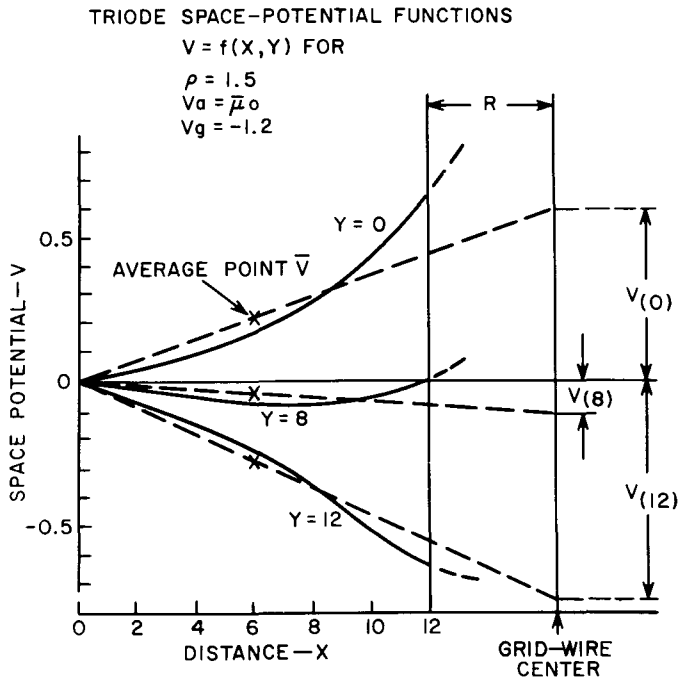


Figure 25. Electrostatic Space-Potential Functions of a Triode

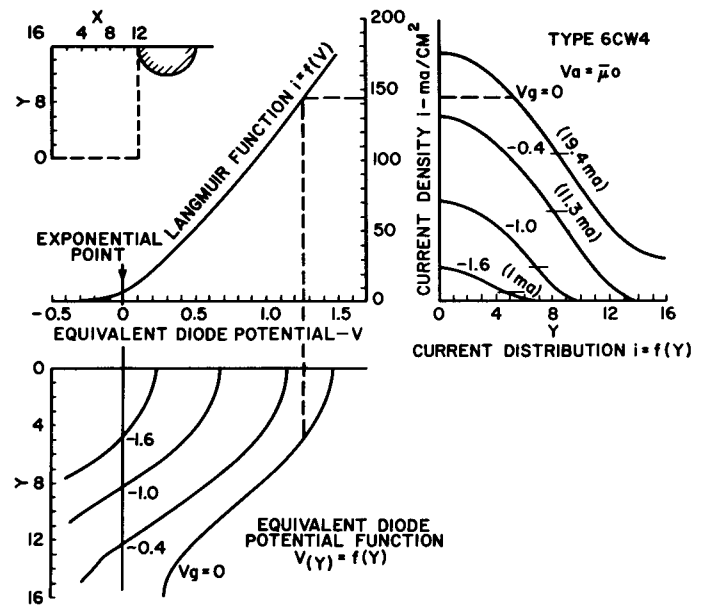


Figure 26. Potential and Current Distribution for Four Grid Potentials (V_g) of a Nuvistor Triode at $V_a = \bar{\mu}_o$

an appraisal of electron trajectories, the reader may rotate the illustration 180 degrees. The space-charge "hill" in front of the cathode is built up by the large number of electrons which have insufficient initial velocities to overcome the retarding fields to the potential minimum or grid plane. The constant-potential contours marked -0.1 and -0.35 on the horizontal planes V = 0 indicate the potential barriers and boundaries of current flow, i. e., points at which the computed current has decreased to 2 per cent of the highest current obtained at the center between grid wires (y = 0). The areas A_o obtained by the assumption of an independent parallel current flow in the y-sections are clearly not sharp limits outside of which electrons from the cathode are turned back. The shaded vertical sections at various distances (x) from the cathode show a spread of the positive V-functions towards the cathode. The effective area A_e of current flow is therefore larger than A_o. The correction factor A_e/A_o for the effective cathode area is clearly a function of the potential modulation in the grid plane. It can be determined by calculation of electron paths or experimentally by comparing the computed currents with currents measured on carefully constructed triodes duplicating the constants used in the

calculations. The correction becomes negligible when the amplitude of the positive portion of the potential modulation is high and when A_0 extends over more than one half the cathode plane (y > 8). For low values of A_0 and low positive potentials in the center between grid wires, however, the number of electrons "funneled" into the area A_0 can represent a substantial percentage of the small computed current, as indicated by the large correction factor $A_e/A_0 = 2$. A set of electron trajectories computed for a similar operating condition is shown in Fig. 29. Because an initial velocity 0.2 volt greater than the cathode-surface potential of 0.65 volt is assumed, no electrons can pass the -0.2-volt potential contours. The relative current distribution is given by the ratio ψ/ψ_0 of the angles at the cathode including all trajectories passing through the grid plane. The inset shows a comparison with the current distribution i/i_0 computed without trajectories. The current computed from the trajectories is larger by the factor $\bar{I}_{Tr}/\bar{I}_c = 1.88$. The agreement with the measured current increase (a factor $A/A_0 = 2$) is a good indication that the discrepancies in the exponential region are caused to a large extent by the assumption of a laminar current flow.

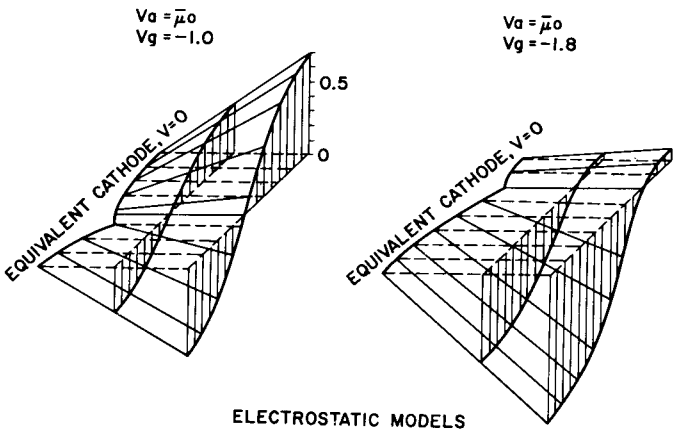


Figure 27. Electrostatic Models Used for Calculation of Equivalent Diode Potentials of a Triode

In view of the good correlation between hand-calculated and measured currents for a large range of practical operating conditions, it was decided to dispense with a second approximation requiring the tracing of electron trajectories in the space-charge-corrected field and set up a computer program for machine calculation of triode (V, I)-characteristics and their derivatives gm, r_p, and μ as outlined above. To obtain program flexibility, and to cover a wide range of parameter variations, it became necessary to combine and transfer earlier partial programs to a large computer (such as the RCA-601 or IBM-7090) having sufficient storage capacity for tables and data and high-speed arithmetic circuits to reduce machine operating costs to attractive values. The following brief description of the machine program 341.07 set up for triode calculations illustrates the large number of operations and relative complexity required for the first-order solution of the problem discussed above. The program description is followed by a discussion of results.

THE COMPUTER PROGRAM

The computer program 341.07 is a numerical analysis program requiring floating-point and high-speed-arithmetic circuits for its economical performance. It includes instructions for calculating the electrostatic potential fields of a triode, and a generalized table of Langmuir relations computed for this purpose. The complete program contains approximately 5300 machine instructions and requires more than 5200 registers for storing tables and incidental information. In addition to performing the calculation of (V, I) characteristics and derivatives for parallel-plane triodes by the method outlined in the preceding section for all practical parameter values, the program contains transformation routines for calculating cylindrical structures, routines for comparing a number of designs, and routines for calculating non-uniform structures by segmentation. Efforts have been made to keep the input simple, to keep the output understandable, and to provide for ease of operation by computer operators.

The program control codes and design data are entered into the computer core memory by means of a program tape and two punched cards per tube design. The data form to be filled out by the engineer is shown in Fig. 30.

A. THE MACHINE PROGRAM FOR PARALLEL-PLANE TRIODES

The sequence of operations carried out by the machine follows.

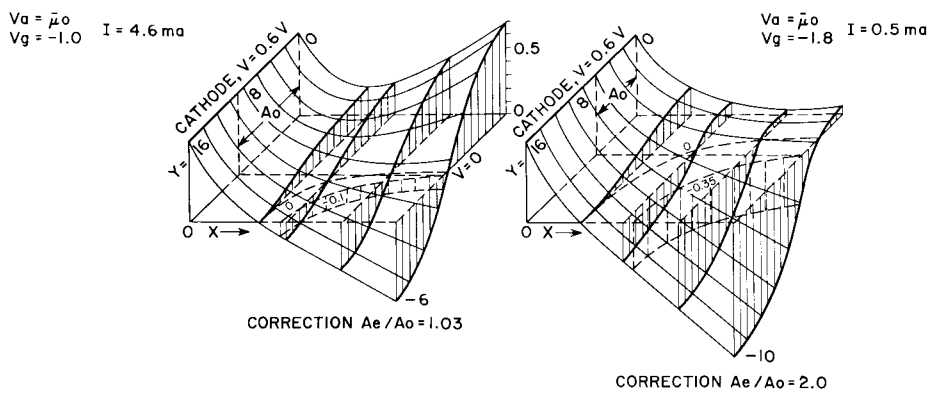


Figure 28. Potential Fields with Space Charge in the Cathode-Grid Space of a Triode Constructed from Fig. 27 and Fig. 25

Calculation of Electrostatic Potential Tables (Sub-Program)

The input data are the grid-cathode distance d_1 (x-dimension), the grid pitch (1/T.P.I.) (y-dimension), the grid-plate distance d_2, and the grid-wire diameter 2R. (The z-dimension of the 3-dimensional spaces is uniform.)

The machine normalizes the dimensions using the grid-wire radius (R) as a unit, and solves the simultaneous equations set up for a system of unipole and multipole line charges located at the grid-wire centers and their image points on opposite sides of the anode and cathode planes to obtain constant potentials at 10 points on the grid-wire diameter (indicated in Fig. 21). With these line charges, the machine calculates the electrostatic unit potentials $v_{g(1)}$ and $v_{a(1)}$ for a matrix of 204 points in the grid-cathode space (x, y) (compare Table V). These unit potentials are calculated for two equivalent grid-cathode distances d_1' related to the real distance d_1 by the ratios $\rho_1 = d_1/d_1' = 1.5$ and $\rho_2 = 1.8$.

The output data from this part of the program are four potential tables (816 potential values) and the electrostatic amplification factors $\bar{\mu}_0$ calculated for the two distance ratios. The data are stored for use in the main program. Machine time for this sub-program is 12 seconds.

Calculation of (V, I) Characteristics (Main Program)

The input data are:

1. The potential tables and the value $\bar{\mu}_0$ for ρ_1 from the electrostatic field program.

2. A fixed table for space-charge functions η, $\xi = f(I_s/I)$ required for calculation of the V, I relation to parallel-plane diodes (modified Langmuir relations).

3. A fixed table of values for the function $\rho_r = f(v_r)$, Fig. 24.

4. The operating parameters specified for the particular case consisting of:

   a. The cathode temperature T (deg K) and work-function ψ_c.
   b. The cathode area A in square centimeters.
   c. The plate-voltage and grid-voltage values specifying points in the characteristics for which plate current and derivatives (gm, r_p, μ) are desired.

The machine calculation proceeds in the following manner (not necessarily in the exact sequence given):

1. The saturated emission current I_s is calculated from the equation $I_s = 120\ T^2 e^{-(\psi_c\ 11600/T)}$. The value $\psi_c = 1.6$ volts is used unless specified. (This

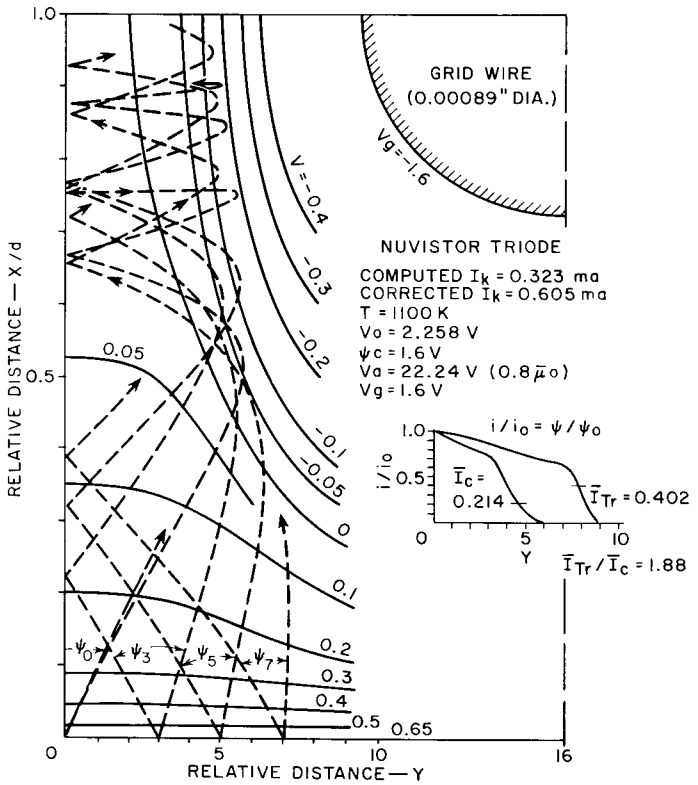


Figure 29. Limiting Electron Trajectories and Corrected Current Distribution (ψ/ψ_0) for a Triode Operating Condition near Current Cutoff

calculation results in $I_S \simeq 7$ amperes per square centimeter for T = 1100K; I_S = 1 ampere per square centimeter for T = 1000K, etc.)

2. The current I_o and the corresponding minimum potential v_{m_o} at the exponential point (dv/dx = 0, V = 0) are calculated for the real grid-cathode distance (d_1) from the Langmuir relations and stored to establish the reference values ξ_o and I_S/I_o in the Langmuir table and the barrier potential $V_o = (V_{m_o} + \psi_c)$ between cathode and grid plane, which is then at the reference potential V = 0. (See Fig. 23.)

3. The ρ_r -function (Fig. 24) is multiplied by scale factors computed from Eq. (36a).

4. Starting with a low plate potential (V_a = 0, for example) and the first grid-potential value (e.g., V_g = +0.6 volt), the machine calculates 12 values of the space-potential function $v_y = f(x)$ with Eq. (32) in the first elemental section of the cathode-grid space at y = 0 (see Fig. 21). It then averages this function, extrapolates, and stores an equivalent diode-potential $V_{(y)}$ for the grid-plane distance d_1' as illustrated by Fig. 25. The calculation is made with the unit potential values $v_{g(1)}$ and $v_{a(1)}$ from the potential tables for the distance ratio ρ_1 = 1.5. The calculation is repeated with the unit potentials for the ratio ρ_2 = 1.8, and a corrected equivalent diode potential is interpolated from the two values $V_{(y)}$ and the stored function $\rho = f(V)$ (Fig. 24). Having determined the equivalent diode-potential $V_{(1)}$ for the first elemental section, the machine calculates the current density $i_{(y)}$ from the stored table of space-charge functions and reference values computed in step 2. This process is repeated for the remaining 16 sections (y) of the cathode-grid space to provide the potential and current distributions $v_d' = f(y)$ and $i = f(y)$ illustrated in Fig. 26 for the particular electrode potentials (V_g, V_a). The total cathode current I_k is calculated by multiplying the mean value of the current-density function i = f(y) by the cathode area.

5. The calculations described in step 4 require less than 0.2 second and are repeated for all specified grid- and plate-voltage points in the tube characteristic. The total current values are stored in matrix form in the core memory. Tape storage of the corresponding distribution functions i = f(y) and V = f(y) for print-out is optional.

6. The derivatives are then calculated from increments in the stored current matrix ($gm = \Delta I / \Delta V_g$, $r_p = \Delta V_a / \Delta I$, and $\bar{\mu} = gm\ r_p$) and stored on the output tape for print-out.

The total machine time for the calculation of four 140-point tables (including the time for the field calculation) is approximately 34 seconds. When a print-out of the potential and current distribution functions for each of the operating points is desired, the machine-time is increased by a few seconds consumed in transferring the information onto the output tape. The final reduction of the taped information to typed data sheets is done by low-cost secondary equipment, as illustrated by Figs. 31A to 31E. A page of the optional information showing the distribution function V, I = f(y) for each operating condition is reproduced in Fig. 31F.

Computed Data

The computed current (Fig. 31B) is the cathode current $I_k = I_b + I_c$ given in milliamperes for the specified cathode area. (Print-out of current densities is optional.)

Computed Potentials (V_g, V_a) and External Voltages (E_c, E_b). The electrode surface potentials V_g, V_a used in the calculations are measured from the exponential point V = 0, dv/dx = 0. The cathode has, therefore, the following positive potential:

$$V_o = \psi_c + V_{m(o)} \qquad (37)$$

where ψ_c is the cathode work-function and $V_{m(o)}$ the

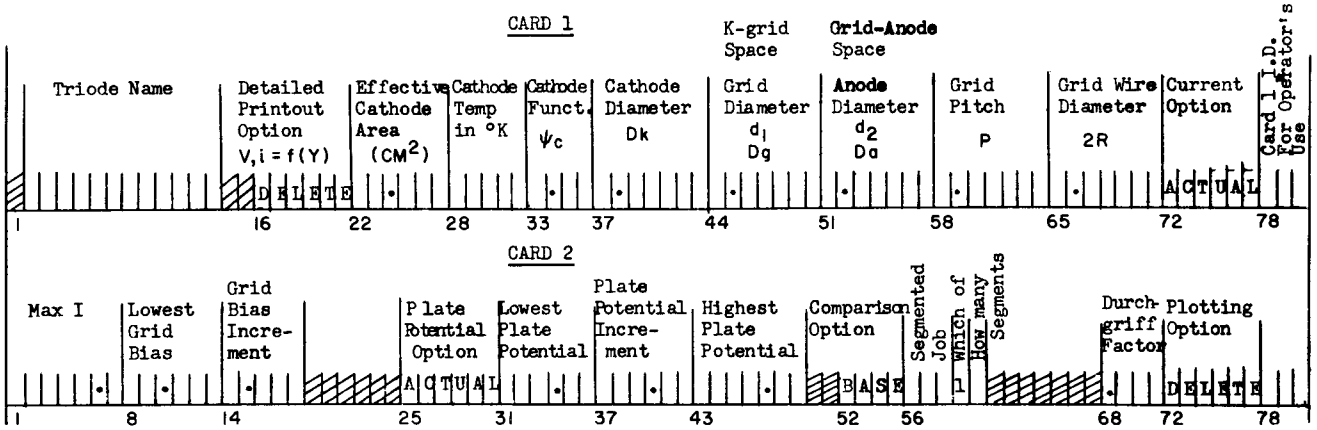


Figure 30. Data Form of Input Instructions for the Machine Calculation of Triode Characteristics

62K80A60G.89

THE INPUTS ARE AS FOLLOWS

HOT CATHODE DIAMETER = 0.06190 IN., GRID DIAMETER = 0.06699 IN., PLATE DIAMETER = 0.08000 IN.

PITCH = 0.00350 IN., WIRE DIAMTER = 0.00089 IN.

THE CATHODE AREA = 0.1850 SQ CM, CATHODE TEMPER ATURE = 1100 K

THE CATHODE WORK FUNCTION = 1.60

THE COMPUTED VALUES ARE AS FOLLOWS

EMISSION CURRENT (I-S) = 6.937 AMPS/SQ CM.

I-INFINITY = 6.948 MILLI-AMPS/SQ CM., I-ZERO = 6.711 MILLI-AMPS/SQ CM.

V-ZERO = 2.25819 MU-ZERO FOR D/D IS 1.5 = 27.797 MU-ZERO FOR D/D IS 1.8 = 28.041

THE CORRECTED MU = 33.012

Figure 31A. Computer-Printed Data Matrices for a Triode: (A) Title Page

62K80A60G.89

CURRENT IS GIVEN HERE IN MILLI-AMPS

BIAS	BEGINNING PLATE = -0. -0.	0.200	0.400	0.600	0.800	1.000	1.200	1.400	1.600	1.800	2.000	2.200
0.40	7.505	11.321	15.544	20.115	0.	0.	0.	0.	0.	0.	0.	0.
0.20	4.096	7.514	11.410	15.694	20.311	0.	0.	0.	0.	0.	0.	0.
0.	1.242	4.135	7.635	11.594	15.928	20.593	0.	0.	0.	0.	0.	0.
-0.20	0.182	1.625	4.429	7.906	11.857	16.252	20.951	0.	0.	0.	0.	0.
-0.40	0.	0.476	2.392	5.223	8.650	12.535	16.813	21.417	0.	0.	0.	0.
-0.60	0.	0.100	1.097	3.281	6.185	9.629	13.507	17.773	22.365	0.	0.	0.
-0.80	0.	0.	0.379	1.882	4.257	7.258	10.752	14.643	18.910	23.512	0.	0.
-1.00	0.	0.	0.099	0.908	2.736	5.293	8.394	11.925	15.854	20.159	24.752	0.
-1.20	0.	0.	0.	0.342	1.611	3.684	6.384	9.571	13.139	17.148	21.446	26.107
-1.40	0.	0.	0.	0.123	0.809	2.397	4.690	7.517	10.779	14.442	18.496	22.839
-1.60	0.	0.	0.	0.	0.323	1.426	3.303	5.751	8.687	12.038	15.773	19.875
-1.80	0.	0.	0.	0.	0.156	0.738	2.168	4.248	6.855	9.892	13.337	17.156
-2.00	0.	0.	0.	0.	0.	0.311	1.310	2.981	5.255	8.001	11.131	14.657
-2.20	0.	0.	0.	0.	0.	0.176	0.688	1.980	3.900	6.323	9.176	12.424
-2.40	0.	0.	0.	0.	0.	0.	0.304	1.227	2.769	4.886	7.429	10.391
-2.60	0.	0.	0.	0.	0.	0.	0.183	0.653	1.825	3.602	5.903	8.574
-2.80	0.	0.	0.	0.	0.	0.	0.	0.299	1.146	2.578	4.538	6.951
-3.00	0.	0.	0.	0.	0.	0.	0.	0.183	0.628	1.725	3.404	5.524
-3.20	0.	0.	0.	0.	0.	0.	0.	0.	0.296	1.072	2.391	4.260
-3.40	0.	0.	0.	0.	0.	0.	0.	0.	0.183	0.617	1.641	3.209
-3.60	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.296	1.023	2.271
-3.80	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.182	0.609	1.563
-4.00	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.297	0.986
-4.20	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.180	0.602
-4.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.300
-4.60	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.180

Figure 31(B). Currents I_k (Zeros are printed for values not calculated)

62K80A60G.89

GM IS GIVEN HERE IN MICRO-MHOS

THE BEGINNING PLATE WAS -0.

BIAS	-0.	0.200	0.400	0.600	0.800	1.000	1.200	1.400	1.600	1.800	2.000	2.200
0.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.20	15659.2	17966.1	19771.6	21301.5	0.	0.	0.	0.	0.	0.	0.	0.
0.	9785.4	14722.5	17452.5	19468.5	21134.5	0.	0.	0.	0.	0.	0.	0.
-0.20	0.	9148.2	13107.0	15929.1	18194.1	20145.3	0.	0.	0.	0.	0.	0.
-0.40	0.	3811.6	8328.5	11564.4	14178.4	16557.7	18608.1	0.	0.	0.	0.	0.
-0.60	0.	0.	5034.3	8351.9	10982.0	13191.7	15152.7	16933.9	0.	0.	0.	0.
-0.80	0.	0.	2496.0	5931.6	8624.4	10840.0	12783.4	14620.0	16277.7	0.	0.	0.
-1.00	0.	0.	0.	3848.5	6616.3	8935.3	10921.8	12679.9	14427.6	15908.1	0.	0.
-1.20	0.	0.	0.	1962.9	4816.2	7238.3	9261.1	11020.8	12686.8	14293.7	15639.9	0.
-1.40	0.	0.	0.	0.	3218.2	5644.8	7700.2	9551.9	11130.0	12776.4	14183.3	15579.9
-1.60	0.	0.	0.	0.	1633.2	4148.0	6304.6	8172.2	9810.4	11373.3	12896.9	14207.5
-1.80	0.	0.	0.	0.	0.	2786.9	4983.1	6924.1	8580.8	10091.1	11605.6	13045.9
-2.00	0.	0.	0.	0.	0.	1406.3	3699.1	5670.6	7387.6	8922.5	10404.3	11829.2
-2.20	0.	0.	0.	0.	0.	0.	2515.7	4384.8	6214.9	7788.0	9253.8	10666.0
-2.40	0.	0.	0.	0.	0.	0.	1264.1	3316.1	5185.4	6803.2	8182.2	9625.0
-2.60	0.	0.	0.	0.	0.	0.	0.	2320.6	4057.0	5769.5	7226.7	8597.9
-2.80	0.	0.	0.	0.	0.	0.	0.	1176.6	2994.0	4693.2	6247.8	7625.3
-3.00	0.	0.	0.	0.	0.	0.	0.	0.	2124.7	3766.2	5367.9	6727.7
-3.20	0.	0.	0.	0.	0.	0.	0.	0.	1112.7	2768.6	4407.5	5787.2
-3.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	1939.7	3420.3	4972.8
-3.60	0.	0.	0.	0.	0.	0.	0.	0.	0.	1088.8	2578.4	4116.2
-3.80	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	1814.3	3212.1
-4.00	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	1072.7	2401.4
-4.20	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	1715.2
-4.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	1054.3
-4.60	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.

Figure 31(C). Transconductances $g_{m(k)}$ (Zeros are printed for values not calculated)

minimum potential (Fig. 23). The value of the potential barrier V_o is printed on the title sheet (Fig. 31A) for each particular case because it is a function of cathode temperature and distance (d_1) and must be known to establish the externally applied voltages E_c and E_b; these voltages are given by

$$E_c = V_g + (\psi_g - V_o) \tag{38}$$

and

$$E_b = V_a + (\psi_a - V_o) \tag{39a}$$

where ψ_g and ψ_a are the average work-functions of the grid or anode surfaces, respectively. The plate potential V_a may be printed out in actual volts or in normalized units $V_{a(n)} = V_a / \bar{\mu}_o$ (option). For normalized units

$$E_b = \bar{\mu}_o V_{a(n)} + (\psi_a - V_o) \tag{39b}$$

The work function ψ_g of the grid cannot be predicted accurately, because it can be changed by processing and aging schedules by as much as 1 volt. It is generally in the order of 2.25 ± 0.5 volt.

The grid voltage for the example (Fig. 31A) is thus $E_c = V_g + 0.06 \pm 0.5$ volt. The additive factor for a given sample tube can be determined accurately by comparison of calculated and measured data, as will be discussed later. It may also be determined with fair accuracy from the difference $V_g - E_c$ by measurement of E_c at a small grid-current value, as follows:

The grid potential V_g is negative for small grid currents. Because of the negative potential, the flux lines are divergent under the grid wires and the grid collects current from a cathode area smaller than the shadow area A_g of the grid. If it is assumed that this fraction is $k = 1/3$ for $E_b = \bar{\mu}_o$ volt, the grid-current density I_g may be approximated by

$$I_g = I_c / K A_g \simeq 3 I_c / A_g \tag{40}$$

For a grid current $I_c = 0.5$ microampere and a shadow area $A_g = 0.3\ A_k = 0.3 \times 0.185$ square centimeter, for example, the grid-current density is $I_g = 27 \times 10^{-6}$ ampere per square centimeter.

62K80A600.89

RP IS GIVEN HERE IN OHMS

THE BEGINNING PLATE WAS -0.

BIAS	-0.	0.200	0.400	0.600	0.800	1.000	1.200	1.400	1.600	1.800	2.000	2.200
0.40	0.	49.8	45.5	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.20	0.	54.7	48.9	44.9	0.	0.	0.	0.	0.	0.	0.	0.
0.	0.	62.6	53.6	48.2	44.5	0.	0.	0.	0.	0.	0.	0.
-0.20	0.	94.2	63.7	53.9	47.9	44.0	0.	0.	0.	0.	0.	0.
-0.40	0.	0.	84.3	63.9	54.7	49.0	45.0	0.	0.	0.	0.	0.
-0.60	0.	0.	125.8	78.6	63.0	54.6	49.1	45.2	0.	0.	0.	0.
-0.80	0.	0.	0.	103.1	74.4	61.6	54.2	49.0	45.1	0.	0.	0.
-1.00	0.	0.	0.	151.7	91.2	70.7	60.3	53.6	48.6	45.0	0.	0.
-1.20	0.	0.	0.	0.	119.7	83.8	67.9	59.2	52.8	48.2	44.6	0.
-1.40	0.	0.	0.	0.	175.9	103.1	78.1	65.7	57.8	51.8	47.6	0.
-1.60	0.	0.	0.	0.	0.	134.2	92.5	74.3	63.6	56.5	51.0	0.
-1.80	0.	0.	0.	0.	0.	198.8	114.0	85.3	70.9	61.7	55.1	0.
-2.00	0.	0.	0.	0.	0.	0.	149.8	101.4	79.7	68.1	60.1	0.
-2.20	0.	0.	0.	0.	0.	0.	221.7	124.6	92.1	75.8	65.6	0.
-2.40	0.	0.	0.	0.	0.	0.	0.	162.3	109.3	85.8	72.7	0.
-2.60	0.	0.	0.	0.	0.	0.	0.	243.5	135.7	98.1	80.5	0.
-2.80	0.	0.	0.	0.	0.	0.	0.	0.	175.5	117.9	91.5	0.
-3.00	0.	0.	0.	0.	0.	0.	0.	0.	259.4	144.1	105.3	0.
-3.20	0.	0.	0.	0.	0.	0.	0.	0.	0.	190.9	125.5	0.
-3.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	274.4	154.3	0.
-3.60	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	202.5	0.
-3.80	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	289.7	0.
-4.00	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
-4.20	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
-4.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
-4.60	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.

Figure 31(D). Plate Resistances $r_p/\bar{\mu}_o$ (Zeros are printed for values not calculated)

The negative grid potential V_g with respect to the exponential point (V = 0) is given by

$$V_g = - \frac{2.3T}{11600} \log I_o/I_g \qquad (41)$$

For the above example and the values given on the data page Fig. 31A, the negative grid potential for a grid current of 0.5 microampere is V_g = -0.575 volt. If k is assumed to be unity, the grid potential V_g = -0.470 volt. The value computed with k = 1/3 is more accurate because k becomes unity only for a parallel current flow, which would require a positive grid voltage near space potential.

The anode work function ψ_a is generally somewhat higher than the grid work function, i.e., $\psi_a \simeq 3 \pm 0.5$ volt. The potentials V_a in Figs. 31B to 31F may, therefore, correspond to slightly higher positive external plate voltages $E_b = V_a + 0.81 \pm 0.5$ volt.

When the dimensions of the sample tube differ from those of the model, the plate-voltage scale is obtained by multiplication of the normalized scale ($V_{a(n)}$) by the value $\bar{\mu}_o$ of the sample tube. A graphic solution for the electrostatic value $\bar{\mu}_o$ of the actual tube will be discussed later.

The Derivatives (gm, r_p, $\bar{\mu}$). The derivatives are calculated with respect to the total cathode current from increments of the current I printed out in Fig. 31B. The transconductance is computed from the increments $\pm \Delta I$ for adjacent grid-voltage increments $\pm \Delta V_g$ at constant plate voltage. Because the first and last rows in Fig. 31C do not give values, the machine indicates a zero symbol. The r_p values are computed from increments $\pm \Delta I$ for adjacent plate-voltage steps $\pm \Delta V_a$ at constant grid voltage; in this case the first and last columns do not give values. The values in Fig. 31D are given in normalized form ($r_p/\bar{\mu}_o$). Because the $\bar{\mu}$ values are the products $\bar{\mu}$ = gm r_p, zero values are given for all border values. The values in Fig. 31E are the normalized ratios $\bar{\mu}/\bar{\mu}_o$.

The derivatives so obtained may be somewhat in error when the increments $\pm \Delta I$ are very unbalanced, i.e., near current cutoff where the curvature of sharp-cutoff characteristics departs considerably from a low-

62K80A60G.89

AMPLIFICATION FACTOR OR MU

BIAS	BEGINNING PLATE = -0. -0.	0.200	0.400	0.600	0.800	1.000	1.200	1.400	1.600	1.800	2.000	2.200
0.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0.20	0.	1.0	1.0	1.0	0.	0.	0.	0.	0.	0.	0.	0.
0.	0.	0.9	0.9	0.9	0.9	0.	0.	0.	0.	0.	0.	0.
-0.20	0.	0.9	0.8	0.9	0.9	0.9	0.	0.	0.	0.	0.	0.
-0.40	0.	0.	0.7	0.7	0.8	0.8	0.8	0.	0.	0.	0.	0.
-0.60	0.	0.	0.6	0.7	0.7	0.7	0.7	0.8	0.	0.	0.	0.
-0.80	0.	0.	0.	0.6	0.6	0.7	0.7	0.7	0.7	0.	0.	0.
-1.00	0.	0.	0.	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.	0.
-1.20	0.	0.	0.	0.	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.
-1.40	0.	0.	0.	0.	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.
-1.60	0.	0.	0.	0.	0.	0.6	0.6	0.6	0.6	0.6	0.7	0.
-1.80	0.	0.	0.	0.	0.	0.6	0.6	0.6	0.6	0.6	0.6	0.
-2.00	0.	0.	0.	0.	0.	0.	0.6	0.6	0.6	0.6	0.6	0.
-2.20	0.	0.	0.	0.	0.	0.	0.6	0.5	0.6	0.6	0.6	0.
-2.40	0.	0.	0.	0.	0.	0.	0.	0.5	0.6	0.6	0.6	0.
-2.60	0.	0.	0.	0.	0.	0.	0.	0.6	0.6	0.6	0.6	0.
-2.80	0.	0.	0.	0.	0.	0.	0.	0.	0.5	0.6	0.6	0.
-3.00	0.	0.	0.	0.	0.	0.	0.	0.	0.6	0.5	0.6	0.
-3.20	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.5	0.6	0.
-3.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.5	0.5	0.
-3.60	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.5	0.
-3.80	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.5	0.
-4.00	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
-4.20	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
-4.40	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
-4.60	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.

Figure 31(E). Amplification Factors $\bar{\mu}/\bar{\mu}_o$ (Zeros are printed for values not calculated)

order parabola. The error becomes negligible if smaller increments ΔV_g are used.

It is evident that a correct choice of potential ranges and increments ΔV_g and ΔV_a is important to center the calculations on the range of interest and avoid unnecessary machine time or test runs.

Input Information

Determination of Potential Scales and Increments ΔV_g and ΔV_a. An estimate of the potential ranges required as input information is readily obtained as follows:

a. Select the highest current value $I_{b_{max}}$ of interest at zero bias. Determine the required diode potential V for the current density $I_{b_{max}}/A$ from the grid-cathode spacing (d_1) and the normalized diode characteristics, Fig. 32.

b. Determine the approximate electrostatic value $\bar{\mu}_o$ from Fig. 33 for the desired plate-grid distance d_2, grid-wire diameter 2R, and grid-pitch (1/T.P.I.).

Calculate the following factor:

$$\alpha = 1 + \frac{1}{\bar{\mu}_o}\left(1 + 1.5\frac{d_2}{d_1}\right) \tag{42}$$

c. The required plate potential for the current $I_{b_{max}}$ at zero bias is

$$\begin{aligned} V_{a(n)} &\simeq \alpha\, V \quad \text{(normalized)} \\ V_a &\simeq \alpha\, V\bar{\mu}_o \text{ (in volts)} \end{aligned} \tag{43}$$

The cutoff bias at this plate potential is

$$V_{g(co)} \geq -(\alpha\, V + 0.4) \tag{44}$$

d. Suitable intervals are $\Delta V_{a(n)} = V_{a(n)}/k$, with k = 5 and $\Delta V_g = V_{g(co)}/k$, where k may have a value between 5 and 15.

Example 1: $I_{b_{max}} = 100\,ma/cm^2$, $d_1 = 0.0025''$, $d_2 = 0.007''$, $\bar{\mu}_o = 70$

From Fig. 32, $V \simeq 0.88$ volt

Y =	0	1	2	3	4	5	6	7
	8	9	10	11	12	13	14	15	16

62K80A60G.89

GRID VOLTS = -3.20, PLATE VOLTS = 50.03 CURRENT PER SQUARE CENTIMETER = 5.7943 MILLI-AMPS.

V =	0.288	0.269	0.210	0.111	-0.003	-0.129	0.	0.
	0.	0.	0.	0.	0.	0.	0.	0.
I =	31.346	29.465	23.977	15.332	6.535	1.726	0.	0.
	0.	0.	0.	0.	0.	0.	0.	0.

GRID VOLTS = -3.40, PLATE VOLTS = 50.03 CURRENT PER SQUARE CENTIMETER = 3.3368 MILLI-AMPS.

V =	0.163	0.143	0.083	-0.001	-0.013	0.	0.	0.
	0.	0.	0.	0.	0.	0.	0.	0.
I =	19.772	18.035	13.029	6.614	5.825	0.	0.	0.
	0.	0.	0.	0.	0.	0.	0.	0.

GRID VOLTS = -3.60, PLATE VOLTS = 50.03 CURRENT PER SQUARE CENTIMETER = 1.6004 MILLI-AMPS.

V =	0.040	0.020	-0.003	-0.011	-0.368	0.	0.	0.
	0.	0.	0.	0.	0.	0.	0.	0.
I =	9.668	8.140	6.505	5.988	0.138	0.	0.	0.
	0.	0.	0.	0.	0.	0.	0.	0.

GRID VOLTS = -3.80, PLATE VOLTS = 50.03 CURRENT PER SQUARE CENTIMETER = 0.9827 MILLI-AMPS.

V =	-0.005	-0.007	-0.011	-0.280	0.	0.	0.	0.
	0.	0.	0.	0.	0.	0.	0.	0.
I =	6.342	6.246	5.955	0.351	0.	0.	0.	0.
	0.	0.	0.	0.	0.	0.	0.	0.

GRID VOLTS = -1.00, PLATE VOLTS = 55.59 CURRENT PER SQUARE CENTIMETER = 133.7954 MILLI-AMPS.

V =	2.236	2.216	2.157	2.060	1.925	1.755	1.551	1.317
	1.059	0.780	0.487	0.179	0.	0.	0.	0.
I =	306.461	303.018	292.896	276.311	253.789	226.051	194.072	159.126
	122.729	86.311	52.036	21.157	0.	0.	0.	0.

GRID VOLTS = -1.20, PLATE VOLTS = 55.59 CURRENT PER SQUARE CENTIMETER = 115.9254 MILLI-AMPS.

V =	2.079	2.059	1.999	1.900	1.762	1.588	1.379	1.142
	0.876	0.588	0.280	-0.127	0.	0.	0.	0.
I =	279.587	276.176	266.028	249.554	227.208	199.745	168.180	134.032
	98.457	63.291	30.589	1.755	0.	0.	0.	0.

GRID VOLTS = -1.40, PLATE VOLTS = 55.59 CURRENT PER SQUARE CENTIMETER = 99.9794 MILLI-AMPS.

V =	1.923	1.903	1.841	1.740	1.599	1.421	1.209	0.966
	0.693	0.395	0.067	0.	0.	0.	0.	0.
I =	253.431	250.016	240.003	223.643	201.504	174.386	143.583	110.195
	75.668	42.193	11.764	0.	0.	0.	0.	0.

Figure 31(F). Detail Print-Out of Potential and Current Distribution (V, I) = f(y) for a Number of Operating Conditions

$$\alpha = 1 + \frac{1}{70}(1 + 4.2) = 1.074$$

hence $V_{a(n)} \simeq 0.94$ and $V_{g(co)} \gtrsim -1.34$ volts

$$\Delta V_{a(n)} \simeq 0.2 \text{ and } \Delta V_g = 0.2 \text{ volt}$$

The plate-voltage range obtainable on one printed page is 11 times $\Delta V_{a(n)} = 2.2\ \bar{\mu}_o$ volts = 154 volts. Shorter or larger ranges may be selected.

Example 2: $I_{b_{max}} = 600$ ma/cm^2, $d_1 = 0.005''$, $d_2 = 0.010''$, $\bar{\mu}_o = 3$

From Fig. 32, $V \simeq 10.2$ volts

$$\alpha = 1 + \frac{1}{3}(1 + 3) = 2.33$$

hence $V_{a(n)} \simeq 24$ and $V_{g(co)} \gtrsim -24.4$

From (d) above: For k = 6, $\Delta V_{a(n)} \simeq 4$ and $\Delta V_g \simeq 4$ volts. The plate-voltage range printed on one page extends to 11 times $\Delta V_{a(n)} = 44\ \bar{\mu}_o$ volts = 132 volts.

Starting Values for V_a and V_g and Current Cutoff. A normal starting value for plate voltage is $V_a = 0$. The starting voltage, however, can be given any desired value. The starting point for the grid voltage can also be given any value, i.e., V_g may start with a positive value, at zero, or with a negative value. The grid bias is automatically increased by the specified increment ΔV_g until the current has decreased to the cutoff value. Although cutoff is usually set to I = 1.5 milliamperes per square centimeter, any other value may be specified.

It is obvious that the machine time is decreased substantially by eliminating the calculation of currents beyond a specified maximum value. Therefore, the program contains an instruction to subtract the value $\Delta V_g = \Delta V_a / \bar{\mu}_o$ from the starting grid-bias value (V_g) in a new plate-potential column when the first current value in the preceding plate-potential column exceeds the specified maximum. The starting bias V_g is thereby reduced progressively, as shown in Fig. 31B.

B. ALTERNATE PATHS AND SUBROUTINES EXTENDING FLEXIBILITY AND SCOPE OF THE PROGRAM

Practical triodes depart more or less in geometry from a perfect parallel-plane structure either by design (cylindrical structures, remote-cutoff designs, etc.) or because of dimensional variations (eccentricity)

introduced by mechanical tolerances. It is therefore desirable to provide alternate paths in the flow of calculations containing subroutines for a conformal transformation of cylindrical structures to equivalent parallel-plane structures and subroutines for the additional segmentation and current addition required in the calculation of non-uniform structures varying in the z-dimension. The calculation of segmented structures further requires a routine for establishing a common reference potential (V = 0) when the exponential points of the segments have different values. This routine is also required to obtain a direct-reading comparison of currents and derivatives in consecutive cases in which the cathode-grid spacing and/or the cathode temperature are not alike.

Subroutine for a Direct Comparison of Computed Characteristics

When a number of triode characteristics are computed in sequence to explore variation of one or several tube parameters, a variation of grid-cathode spacing (d_1) or cathode temperature (T) causes a change of the potential barrier $V_o = (\psi_c + V_{m(o)})$ at the exponential point and hence of the reference potential V = 0 used in the calculations. A direct-reading comparison of currents and gm values requires that the first one of several cases be made the reference case and that the grid and plate potentials $(V_g, V_a)_n$ of all other cases (n) be shifted by an increment $\Delta V_o = V_{o(n)} - V_{o(1)}$ to obtain operating points corresponding to equal external voltages (E_c, E_b). Such operating points are given by

V_g, V_a for the reference or "Base" case

and

$(V_g + \Delta V_o)$, $(V_a + \Delta V_o)$ for all other cases (n) (45)

with

$\Delta V_o = V_{o(n)} - V_{o(1)}$

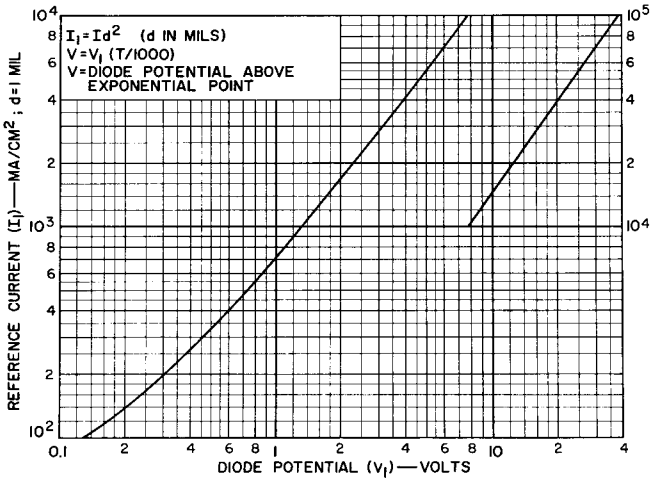

$I_1 = Id^2$ (d IN MILS)
$V = V_1$ (T/1000)
V=DIODE POTENTIAL ABOVE EXPONENTIAL POINT
REFERENCE CURRENT (I_1)—MA/CM^2; d=1 MIL
DIODE POTENTIAL (V_1)—VOLTS

Figure 32. Normalized Diode Characteristic

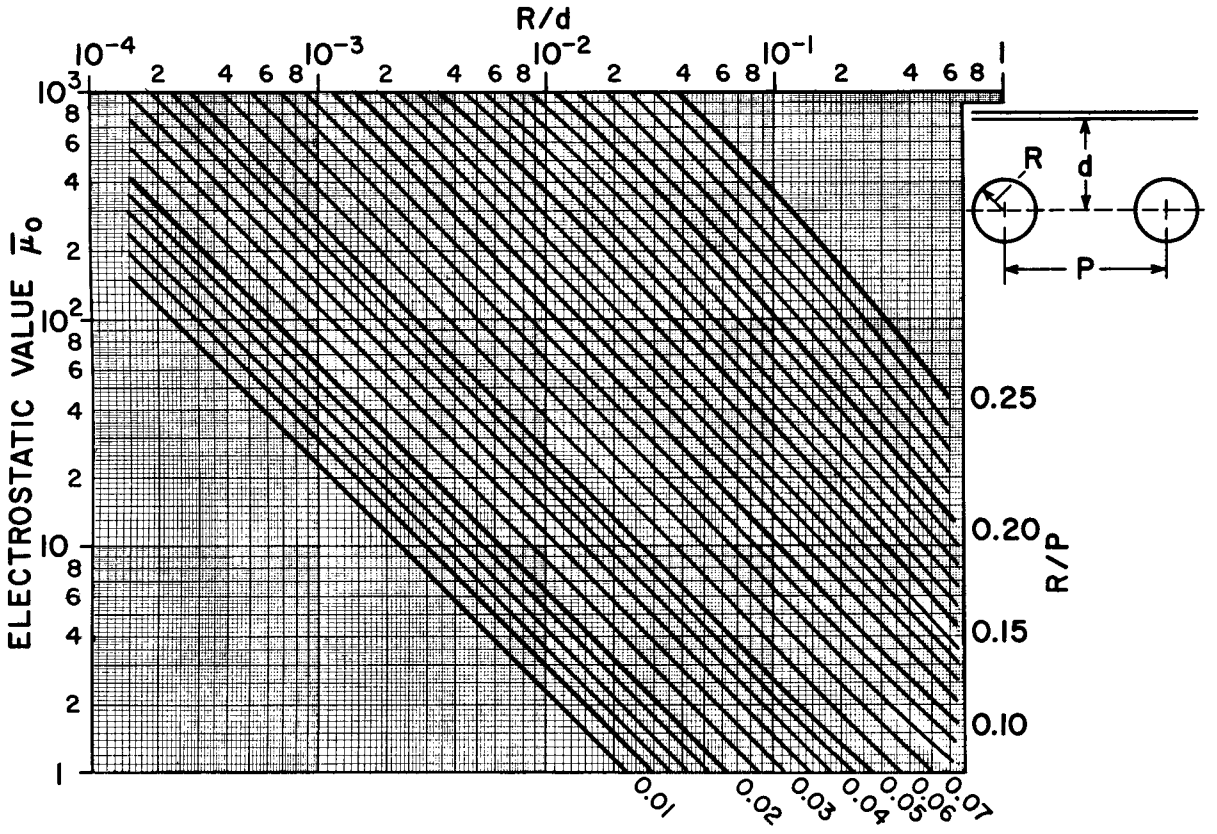


Figure 33. Amplification Factor of Triodes

When instructed to "compare" a series of designs (the word "Base" in the "comparison option" space of Fig. 30 is crossed out) in all but the first data forms the computer will automatically make the "base" design the reference case (its V_0 value) and shift the grid-bias and plate-potential values of consecutive designs (n) according to Eq. (45). The operating points (V_a, V_g) specified for the (V, I) characteristics of a comparison series should therefore be alike or integral multiples, and the plate potentials must be requested in volts (not normalized) to accommodate changes in $\bar{\mu}_0$ values.

The actual values V_0 are given on the title sheets for the different cases, and the currents and derivatives printed in corresponding V_g rows and V_a columns of the data tables refer to identical external voltages (E_c, E_b), although the printed grid and plate-potential values are slightly different according to Eq. (45).

Subroutines for Triodes Having Cylindrical Elements

The elements of a coaxial structure have progressively different surface areas. For a given spacing of two elements, the capacitance of parallel-plane elements of equal area is smaller or larger than the capacitance of cylindrical elements, depending on whether the area A_1 of the inner cylinder (diameter D_1) or the area A_2 of the outer cylinder (diameter D_2) is used for reference. The capacitance ratios are given by

$$C_{par}/C_{coax} = \frac{D_1 \ln (D_2/D_1)}{(D_2-D_1)}, \quad A_1 = \text{reference area} \tag{46}$$

and

$$C_{par}/C_{coax} = \frac{D_2 \ln (D_2/D_1)}{(D_2-D_1)}, \quad A_2 = \text{reference area} \tag{47}$$

For electrostatic-field calculations, a cylindrical triode can be replaced by a parallel-plane triode model having a control grid of equal area, grid pitch (P_0), and wire diameter ($2R_0$), and equal grid-cathode and grid-plate capacitances. The spacings d_1 and d_2 of the equivalent parallel-plane model are therefore obtained by letting the capacitance ratios (Eq. 46 or 47) equal unity.

The grid-cathode distance of the equivalent parallel-plane model used for calculations of the potential functions of a coaxial triode having an internal cathode is thus obtained from Eq. (47):

$$d_1 = (D_g/2) \ln D_g/D_k \tag{48}$$

The grid-anode distance is obtained from Eq. (46):

$$d_2 = (D_g/2) \ln D_p/D_g \tag{49}$$

The computer will perform this transformation when the data card (see Fig. 30) contains three values (cathode diameter, grid diameter, and plate diameter) specifying a coaxial structure, instead of two values (cathode-grid space d_1 and grid-plate space d_2) indicating a parallel-plane structure.

The electrostatic values $\bar{\mu}_0$ and the unit potential functions for the equivalent distance ratios $\rho = d_1/d_1'$ are calculated by the computer from the transformed model dimensions. The calculation of currents, however, requires the numerical value of the actual or equivalent grid-cathode distance. The distance d_1 of the transformed model [Eq. (48)] must therefore be scaled to duplicate the flux density obtained at the location of the equivalent cathode in the space-charge-free coaxial triode model.

It is evident from Fig. 34 that the width Δy and flux density of a section (y) in a parallel-plane diode are constant for a given flux value at all distances (x) from cathode to grid, whereas the flux density in a corresponding pie-shaped section of a coaxial diode is inversely proportional to the radius length, having its lowest value on the mean grid diameter D_g used as reference surface for the electrostatic model. The equivalent cathode in the parallel-plane model is located at the distance $d' = d_1/\rho$. It occurs in the coaxial model on a cylindrical surface having the diameter $D\rho$ which must be made the reference surface. The scaling factor derived from the logarithmic relations is given by

$$D_\rho/D_g = (D_k/D_g)^{1/\rho}$$

The spacing of the equivalent diode for current calculations is, therefore,

$$d_{1(equ)} = d_1(D_k/D_g)^{1/\rho} \tag{50}$$

The computer is instructed to calculate $d_{1(equ)}$ from Eq. (50) for the fixed ratio $\rho = 1.5$, although $d_{1(equ)}$ is actually a function of V in coaxial triodes. The error in the computed current resulting from this simplification does not exceed $\pm$ 5 per cent for $(D_k/D_g) = 0.8$, and increases to $\pm$ 10 per cent for $(D_k/D_g) = 0.65$. The correction for variable ρ values has been omitted in the present program because it is a second-order effect, affecting only a fraction of the current function, and most modern tubes have ratios $(D_k/D_g) \geq 0.8$. Inclusion of the variation $d_{1(equ)} = f(\rho)$ would result in a large increase in machine time.

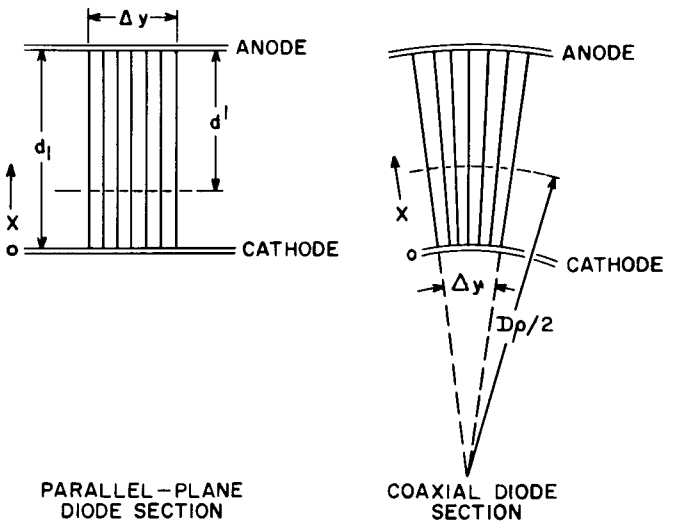


Figure 34. Electrostatic Flux Lines in Parallel-Plane and Coaxial Diodes

Correction for Mesh Grids

A mesh grid can be regarded as two crossed parallel-wire grids located in the same plane (true mesh) or behind one another. Replacement of the mesh-grid triode by an equivalent triode having a parallel-wire grid requires that their electrostatic values $\bar{\mu}_0$ be identical and that the potential field in the grid-cathode space be equivalent. A true mesh having the pitch values P_0 and $P_1 = P_0$ may be replaced by a parallel-wire grid of equal wire diameter (2R) having a finer pitch P_{11} which can be determined from Fig. 35.* When the cross wires (P_1) are not in the same plane but are located on the plate side of the finer-pitch control wires P_0 (for example, the helix winding of a nuvistor grid which is wound over the parallel cage-grid wires), their main effect is a decrease of the unit plate-field intensity as obtained with a larger grid-anode distance.

This type of mesh grid, therefore, can not be replaced by an equivalent parallel-wire grid of finer pitch, which results in higher gm values and a sharper cutoff, but it can be replaced by a μ correction factor and a parallel-wire grid having the actual pitch P_0 of the control wires. The reciprocal $\bar{\mu}$ correction factor or "D factor" ("Durchgriff" Factor) is shown in Fig. 36 as a function of the pitch ratio P_0/P_1 and the ratio $2R/P_0$ of wire diameter to pitch. (A mean value may be used when the wire diameters of P_0 and P_1 are not alike.) The computer uses the corrected value $\bar{\mu}_0/D$ in potential calculations when a D factor is specified in card 2 (Fig. 30), and prints out both the uncorrected and the corrected $\bar{\mu}_0$ values, as shown in Fig. 31A.

Nonuniform Spacings or Parameters (Segmentation Routine)

The potential fields in structures having nonuniform spacings between elements are not uniform in the z-dimension, as in parallel-plane or coaxial structures; i.e., $v = f(x, y, z)$. Small variations Δd can be absorbed by the use of average spacing values; for larger variations, however, it is necessary to extend the segmentation of the grid-cathode field to the z-dimension. A simple case is a parallel-plane structure having different element spacings on opposite sides of a planar cathode. It is obvious that the solution is obtained by two separate triode calculations (for the two halves) and subsequent addition of currents. A similar case is that of a cylindrical structure in which one of the elements is not coaxial.

Assume, for example, a cylindrical triode in which the grid axis has a parallel displacement Δd. The sinusoidal spacing variations $\Delta s \sin \phi$ around the tube axis (z-dimension) can be replaced by a "step" variation. The step levels may be arithmetic mean values of the spacings ($s + \Delta s \sin \phi$) in these segments because the increments of s are generally small compared to s. The tube segments have equal cathode and plate diameters but different cathode areas and grid diameters.

*Based on data given in the 1957 IRE Convention Record, Part 3, pp. 114-119.

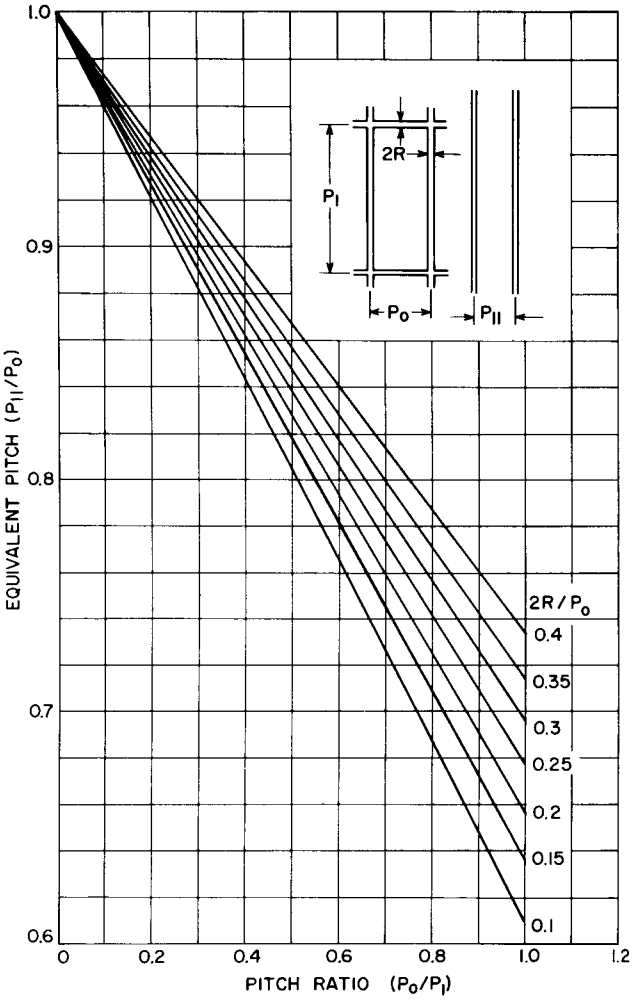


Figure 35. Equivalent Parallel-Wire Pitch (P_{11}) of True-Mesh Grids

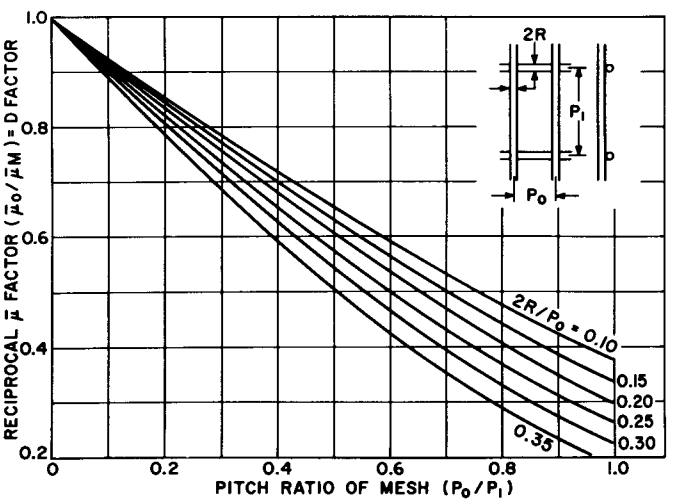


Figure 36. D Factor of Mesh Grids (See text)

The (V, I)-characteristics of the triode segments are computed for direct comparison to include the grid- and plate-potential shifts ΔV_o caused by changes of the grid-cathode distance. The computer can then add currents from the segments directly into one current matrix when instructed that successive specifications are segments of one case (compare Fig. 30, card 2, box labeled "segmented job," etc.).

Although three segments are usually sufficient for a good approximation by discrete steps, the program can accommodate as many segments as desired.

The optional print-out of current-distribution functions for all segments is available for a detailed study of the currents contributed by the segments to the total current sum. Fig. 37 shows the potential and current density distributions obtained by a three-segment calculation of an eccentric coaxial triode for one operating condition (V_a = 80 volts, V_g = -0.55 volt). Assumed is a parallel displacement of the grid axis by Δd = 0.0008 inch, representing a 32 per cent eccentricity in grid-cathode spacing. The current densities in segment 2 are the same as those obtained with a coaxial structure for which a total average plate current I_b = 7.92 milliamperes and a transconductance gm = 12560 micromhos are computed. The eccentricity causes a 15 per cent increase in plate current (I_b = 9.115 milliamperes) and an 8.6 per cent increase in transconductance (gm = 13,665 micromhos) because of the closer spacing and higher current densities in segment 1 (see Fig. 37).

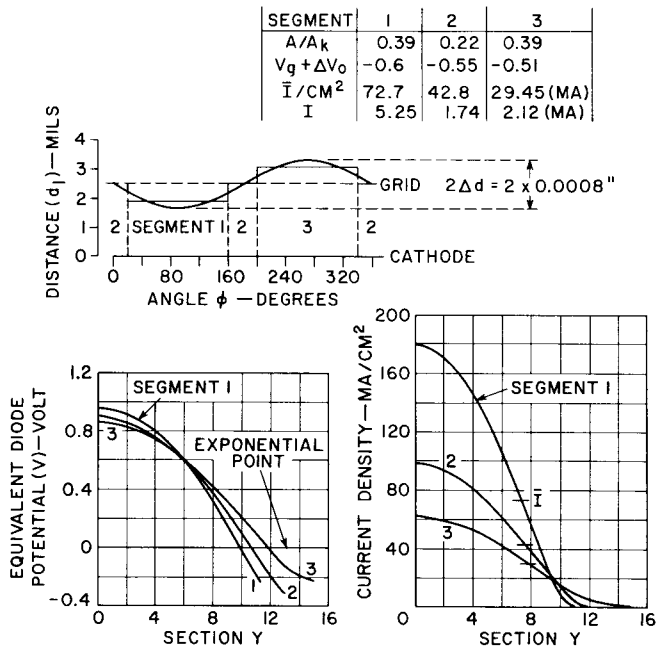

SEGMENT | 1 | 2 | 3
A/Ak | 0.39 | 0.22 | 0.39
Vg+ΔVo | -0.6 | -0.55 | -0.51
Ī/CM² | 72.7 | 42.8 | 29.45 (MA)
I | 5.25 | 1.74 | 2.12 (MA)
DISTANCE (d)—MILS
GRID 2Δd = 2 x 0.0008"
SEGMENT 1
CATHODE
ANGLE φ —DEGREES
EQUIVALENT DIODE POTENTIAL (V)—VOLT
SEGMENT I
EXPONENTIAL POINT
SECTION Y
CURRENT DENSITY—MA/CM²
SECTION Y

Figure 37. Spacing Variation (top), Potential Functions V = f(y), and Current Distributions I = f(y) of an Eccentric Triode

It is apparent that the segmentation routine can be applied to compute the characteristics of structures having several grid-pitch values (remote-cutoff tubes) or other dimensional variations, as well as emitter nonuniformities such as nonuniform work functions or temperatures.

CORRELATION OF COMPUTED AND MEASURED DATA

Measurement of (V, I) Characteristics

The parameters of actual tubes are generally not a precise duplicate of the mathematical model. Cathode temperature, area, electrode spacings, and surface work functions may not be uniform or may deviate from assumed values, and the actual grid may not have exactly the electrostatic field and value μ_o computed for an equivalent parallel-wire grid. The cathode temperature, for example, is a function of cathode current because of the heat-power loss $P_e = V_o I_k + 2Kt/e$ caused by electron evaporation. A constant cathode temperature requires, therefore, an adjustment of heater power according to current; otherwise, the rate of change in current during a measurement must be faster than the thermal time constant of the cathode surface. The dynamic measurement of (V, I) characteristics with a cathode-ray curve tracer and camera equipped with optical graticule projection (eliminating parallax errors) comes close to meeting this requirement, and currents and voltages can be measured with a precision of $\pm$ 1 per cent for all ranges. The operating mode (i.e., the number and value of bias steps and the ac plate-voltage scan) must remain fixed during a measurement to maintain a constant average current and a substantially constant cathode temperature. The current or plate-voltage sensitivity of the amplifiers may be varied to expand the scales (over-scanning). A large number of small grid-bias steps (without slowing down the repetition rate) can be provided by simple circuit modifications to permit interlacing of two sets of bias lines by a double exposure. To obtain 22 grid-voltage steps ΔE_c = 0.2 volt, for example, the step voltage is set to ΔE_c = 0.4 volt and 11 steps for one exposure. A second exposure is then made with the starting voltage shifted by -0.2 volt. A photograph of an interlaced characteristic is shown in Fig. 38.

When the cathode temperature in a high-perveance tube differs by $\Delta T = \pm$ 50 K from the computed value, the potential barrier V_o changes by $\Delta V_o \simeq \pm 0.1$ volt. The change in current can be counteracted by an opposite change of grid voltage $\Delta E_c \mp 0.1$ volt, but not completely for all current values because of a residual difference in the (V, I) characteristic. This difference is illustrated in Fig. 39 for a large temperature change ΔT = 200 K causing a bias change ΔE_c = -0.4 volt.

Analysis of Measured Characteristics

The parameters of sample tubes deviate more or less from the constant values postulated in the calculation of (V, I) characteristics. Grid and plate voltages differ from the computed potentials V_g and V_a by constants depending on processing and aging schedules. Dimensional deviations which occur during assembly or processing may change the perveance, the grid-wire diameter, the electrostatic value $\bar{\mu}_o$, and the current scale; variations in thermal emissivity and work function may

alter the assumed values for cathode temperature and emission. The (V, I) characteristics may also be degenerated by interface or coating resistances.

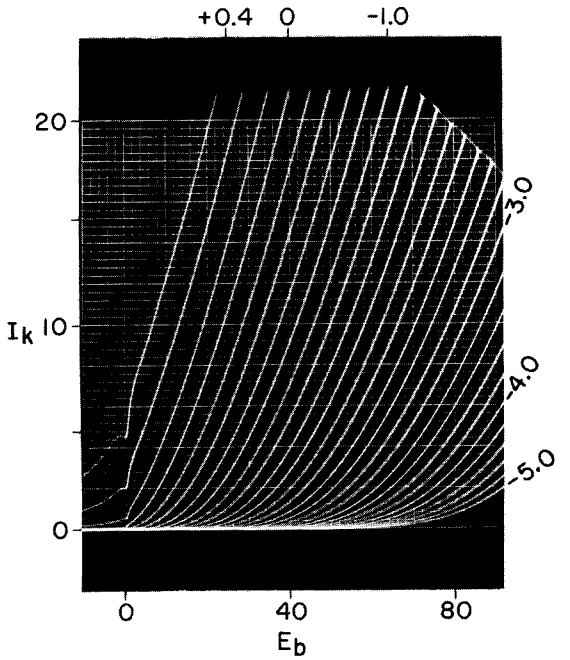


Figure 38. Double-Exposure Photograph of (V, I) Characteristics of a Nuvistor Triode Obtained by Interlacing Grid-Bias Lines

The unknown parameters can be determined or indicated as follows:

The calculated and measured (V, I) characteristics are plotted on separate sheets of log-log coordinate paper, as shown in Fig. 40, for the measured characteristics shown in Fig. 38. When normalized plate potentials and increments are used, points of constant potential $V = (V_g + V_a/\bar{\mu}_o)$ can be connected in the computed characteristic by auxiliary lines (V) for the purpose of accurate grid-voltage interpolation because the V lines intersect the bias lines at constant plate-potential increments $\Delta V_a/\bar{\mu}_o$. The linear plot in Fig. 41 shows that a constant potential line (V) departs considerably from the constant current level indicated by the classic solution, which is very much in error because it assumes uniform potential and current distributions.

The computed characteristic (log-log) is placed over the measured characteristic (log-log), with the plate-potential scale at $V_a/\bar{\mu}_o = 1$ lined up with the plate-voltage scale at the voltage E_b equaling the computed value $\bar{\mu}_o$; the correlation of slope values of the bias-line "fans" is then observed in the two characteristics. A horizontal shift to the right or left may be necessary to obtain the best match of slope values and bias-line spacings, which may not occur on the drawn bias lines but on interpolated lines. A perfect match is generally limited to the grid-current-free range (i. e., $V_g \geq -0.3$ volt) and to currents $I > 1.5\ I_o$ to $3\ I_o$, as shown by the broken lines in Fig. 41. The value $\bar{\mu}_o$ is given by the plate voltage under the potential $V_a = \bar{\mu}_o$, and the average grid work function ψ_g is obtained from the difference $(E_c - V_g)$ and Eq. (38).

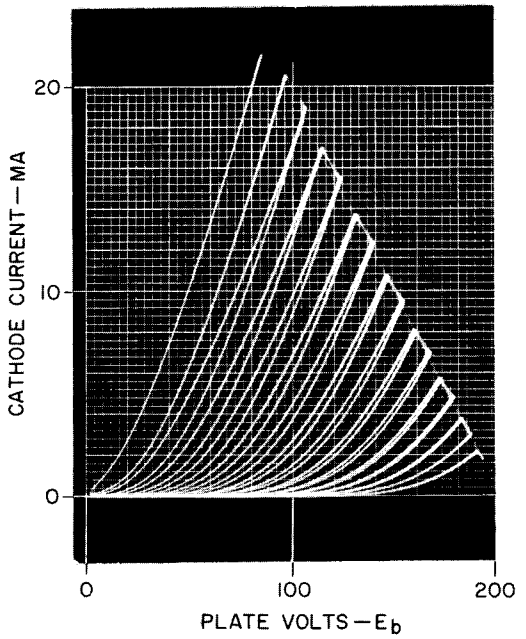


Figure 39. Double-Exposure Photograph Illustrating Change of (V, I) Characteristic of a Close-Spaced Triode (Nuvistor) Caused by a Change of Cathode Temperature $\Delta T = -200$ K. The lower temperature shifts the first zero-bias line to lower currents, corresponding to the third bias line ($\Delta E_c = -0.4$ v) of the higher-temperature characteristic.

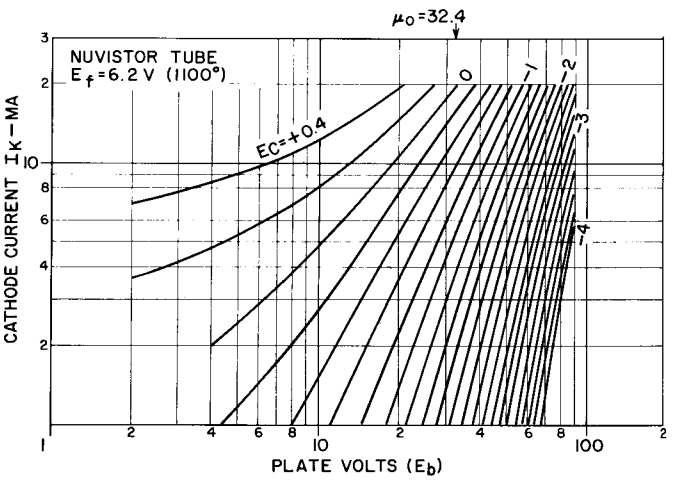


Figure 40. Log-Log Plot of (V, I) Characteristics Shown in Fig. 38

In the event that a horizontal shift does not produce an accurate match, the cathode area (A), the grid-cathode distance (d_1), or the cathode temperature may be in error. A vertical shift multiplies the current scale by a factor, thereby correcting accurately for an error

in cathode area, and in good approximation for small errors in grid-cathode spacing (perveance). An error in cathode temperature is also corrected in first approximation, but a residual curvature error may remain in the grid-bias lines which follow a different power law, as shown in Fig. 39. High coating or interface resistance, as well as pitch or spacing nonuniformities, can also prevent a good match of the computed bias-line curvature because they alter the power law of (V, I) characteristics. Errors of this type are indicated when a rotation of coordinates is required to match computed and measured characteristics.

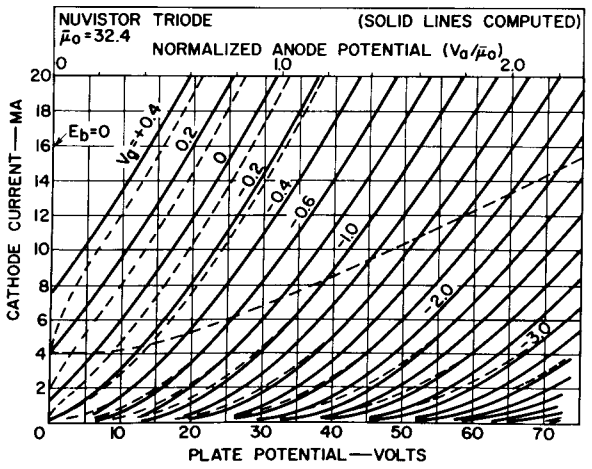


Figure 41. Linear Plot of Computed (V, I) Characteristics (Solid lines) Showing Correlation with Measured Characteristics (Broken lines)

CONCLUSIONS

Comparisons of computed data with values measured on carefully constructed electron tubes have shown excellent agreement of currents and derivatives over a wide range of voltages and currents. The computed current is in error in the exponential region (near current cutoff) because of the assumption of independent current flow in the y-sections, as discussed in connection with Fig. 29. It appears quite possible that a correction for the converging electron flow can be expressed as a simple function of the potential modulation in the grid-cathode space to provide higher accuracy in the cutoff region.

In the grid-current region (V_g more positive than approximately -0.4 volt) the constant grid-voltage lines are shifted to slightly higher anode potentials (ΔV_a = 1 to 3 volts). This error (compare Fig. 41) is caused by neglecting the space charge added by orbiting electrons, secondary emission, or electron reflection on the grid wires. The assumption of a uniform grid-wire potential may also cause deviations at low potentials because the work function (ψ_g) of the wires is reduced only on two-thirds of their surface (visible to the cathode) by deposits of emitter coating evaporated from the cathode, while approximately one third (facing the anode) retains the higher work function of substantially clean metal, as indicated in Fig. 42. The grid field is thus actually a dipole field: $v_g = f(x, y) = (V_g + a)\, v_{g(1)f} + (V_g - b)\, v_{g(1)b}$, where a and b are constants specifying the difference $\Delta\psi_g$ from the mean work function ψ_g for the unit potentials $v_{g(1)f}$ and $v_{g(1)b}$ of the front and back surfaces of the grid wires, respectively. The sum of the unit field components $v_{g(1)f} + v_{g(1)b}$ is equal to the unit field $v_{g(1)}$ for a uniform grid-wire potential (a = b = o).

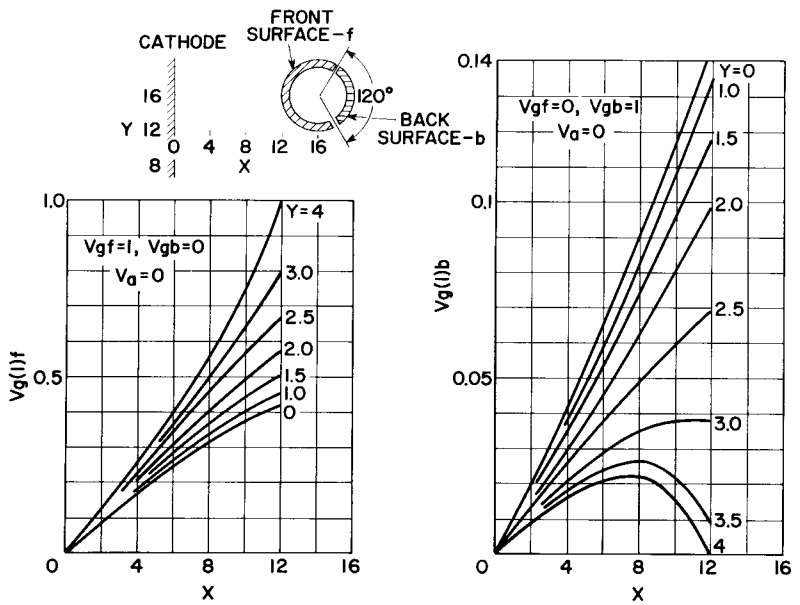


Figure 42. Unit Potential Fields of a "Dipole" Grid in the Model X/Y/C = 4/4/28. The Sum of Both Fields is Equal to the Unit Grid Field Given in Table V.

The "front" field of the grid wires is quite similar to the normal unipotential grid field of a slightly smaller grid wire, while the "back" field of the grid wires has the form of an anode field, as shown in Fig. 42. In a future program, the two unit grid fields may be computed by instructing the computer to assign zero potential to either four or six of the 10 points on the grid-wire surfaces (compare Fig. 21) in solving the simultaneous equations for the line charges at the grid-wire center. The calculation of effective potentials in the plane of second and third grids may also be included for calculation of the cathode current in tetrodes and pentodes as a function of all element potentials.

It can be stated in summary that the accuracy of the machine-computed characteristics is far better than that obtained by other methods of calculation, and the experience gained so far indicates that the machine program is a new analytical tool of great potential for the prediction, analysis, and optimization of electron-tube performance.

ACKNOWLEDGMENTS

The assembly and successful operation of the complete machine program for triode calculations is due to the able programming and painstaking efforts of Hugh Cort, Jr. of the RCA Data Systems Group in Harrison. The instructions for calculation of the electrostatic potential fields of triodes, and a table of space-charge relations based on the Langmuir relations were taken over from two earlier machine programs set up by W. A. Harris of Tube Applications and H. W. Kirk of RCA Data Systems and Services, Electron Tube Division, Harrison. The author wishes to express his appreciation to these men and has profited from their experience and helpful criticism.

REFERENCES

1. Hermann, I. G. and P. S. Wagener, The Oxide-Coated Cathode, Vol. 2, Physics, 1951
2. Van Der Ziel, A., "Extension and Application of Langmuir's Calculations on a Plane Diode with Maxwellian Velocity Distribution of the Electrons," Philips Research Reports 1, pp. 97-118, Jan. 1946
3. Ferris, W. R., "Some Characteristics of Diodes with Oxide-Coated Cathodes," RCA Review, Vol. X, No. 1, March 1949
4. Kleijnen, P. H. J. A., "Extension of Langmuir's (ξ, η) Tables for a Plane Diode with a Maxwellian Distribution of the Electrons," Philips Research Reports 1, pp. 81-96, Jan. 1946
5. Diemer, G. and H. Dijkgraaf, "Langmuir's (ξ, η) Tables for the Exponential Region of the I_a-V_a Characteristic," Philips Research Reports 7, pp. 45-53, 1952

LIST OF SYMBOLS

A	Area of cathode
C	Capacitance
D	Durchgriff = $1/\mu$
d	Grid-to-cathode distance
d'	Equivalent distance
d_m	Distance of potential minimum to cathode
E_e	External field gradient-volts/cm
E_b	Plate voltage (external)
e	Base of natural logarithms = 2.71828
g	Conductance
g_m	Transconductance
i	Current in a triode section (y)
I	Average current
I_o	Current at exponential point ($d_m = d$, $V = 0$)
$\bar{I}_o$	Average current without eccentricity or at zero vibration amplitude
I_s	Saturated emission current
I_r	Current for retarding fields (exponential region)
k	A factor
l	Distance normalizing factor, Eq. (12d)
m	Modulation factor
n	Number of electrons
P_e	Heat power loss caused by emission of electrons, Eq. (4)
P_c	Power loss due to heat conduction from cathode
P_h	Heater power
P_r	Power loss due to heat radiation from cathode
q_e	Electron charge
Q	Total charge
R	Grid-wire radius
T	Temperature, for emitters in deg K
V	Electrode potential
V_a	Anode potential
V_g	Grid potential
V_d	Diode potential between work-function barriers of cathode and anode (Fig. 2)
V_m	Minimum potential, Eq. (14)
V_{m_o}	Minimum potential at exponential point
V_o	Total potential barrier at exponential point, Eq. (8)
V_T	Mean electron-volt velocity, Eq. (2)
$V_{\eta a}$	Accelerating potential, Eq. (15), Fig. 2b
v	A potential in space
v_a	Space potential component due to anode field
v_g	Space potential component due to grid field
$V_{a(1)}$, $v_{g(1)}$	Space potential components caused by unit electrode potentials $V_a = 1$, $V_g = 1$
$\bar{v}$	Average value of space potentials along a specified coordinate (x or y)
x	Distance coordinate normal to cathode surface
y	Distance coordinate parallel to cathode surface
$X=d_1/\rho R$	Cathode to grid-center distance, normalized
Y	Half-distance between grid-wire centers (2Y = grid pitch, normalized)
$C=d_2/R$	Grid-center to anode distance, normalized
α	Control factor, Eq. (26a)
η	Normalized potential, Eq. (13), Figs. 4 and 7
ξ	Normalized distance, Eq. (12)
τ	Electron transit time
ψ	Work function (average emission value for emitters, average potential value for nonemitters)
ρ	Distance ratio d/d', Eq. (25) and Fig. 24
ψ_o	Uniform work function
ψ_e	Effective work function reduced by external field of E_e volts/cm
$\bar{\mu}_o$	Electrostatic average amplification factor
$\bar{\mu}$	Effective electrical average amplification factor
γ	Potential gradient at anode, Eq. (36)

Metals and Metallurgy

C. W. Horsting, et al

Harrison

THE METALLIC STATE

C. W. Horsting

The first section of this article on metallurgy is meant to provide a background of information concerning the metallic state. Of necessity it cannot be more than an introduction; for those who desire to go into greater detail, a list of recommended textbooks, in addition to the quoted literature references, is given at the end of this section.

Great progress has been made in metallurgy since the turn of the century; what used to be an art has developed into a scientifically well-based technology.

In the early nineteen hundreds, metallurgy consisted mainly of an empirical body of knowledge which was applied without much understanding. Testing of mechanical properties was known and used as a test tool for quality; chemical analysis was used to determine and reproduce alloys; thermal expansion was measured; and considerable work was done on microscopic examination of metal structures. Thermal analysis also was a field of investigation, and led to a substantial amount of useful data on phase-equilibrium diagrams of binary-alloy systems. However, no fundamental scientific data were available to explain the observed facts: why could a metal be deformed plastically; what was the mechanism allowing this process to take place; how did work hardening or annealing occur.

ATOMIC STRUCTURE

One of the giant steps in our theoretical approach to fundamental problems was the discovery and application of X-ray diffraction analysis to metals. Von Laue[1], in Germany had suggested that the wave length of X-rays should be just about right for the lattice in crystalline materials to act as a diffraction grating; the subsequent experiments using a nickel sulfide crystal showed that he was right. Soon thereafter, the Braggs[2] in England developed the X-ray analysis technique which established that metals are crystalline substances, notwithstanding the fact that the metallic objects which we observe in daily life seldom show well-defined crystal faces. The internal structure is essentially crystalline. This realization changed the thinking on the subject radically and has been responsible for great progress in the theoretical approach. If metals, like other known crystals, were made up of neatly arranged atom patterns, as the X-ray results indicated, a basis had been found for new explanations of observed facts. Deformation without failure was explained by assuming glide motion between atomic crystal planes. The mysterious matter of solid solutions, the terminal phases in binary equilibrium diagrams which did not show dual structure, but seemed homogeneous under microscopic examination, could now be explained by assuming that atoms of one kind had replaced some of the atoms of the other kind in the crystal lattice. Further, the fact that the lattice spacing, as determined by X-ray, showed a continuous change with the addition of a foreign atom, led to the concept of atomic size as a factor in explaining compatibility in such mixtures, i. e., solubility. This concept led Hume-Rothery[3] in England to the suggestion of a general solubility law which says that mutual solubility is possible when atomic sizes do not differ more than approximately 14 per cent.

Another important result of X-ray analysis was the realization that most metals crystallize in one of three crystal structures: body-centered cubic (B.C.C.), face-centered cubic (F.C.C.), and hexagonal close-packed. Fig. 1 shows their lattice structures.

The space patterns shown in these drawings are called "unit cells." These are by definition the simplest combinations of atoms which, by repetition in three dimensions, can produce the complete, finite space lattice. The Braggs, in their book, use the idea of a wallpaper pattern as a two-dimensional example to illustrate the point.

The basic differences among these structures with respect to the relative location of the atoms in space, manifest themselves, in corresponding differences in physical properties. For instance, while the B.C.C. and F.C.C. lattices are substantially isotropic, the hexagonal lattice has a preferred direction along the hexagonal axis, which leads to specific consequences. Thermal and electrical conductivity, for instance, in

* This article, which comprises nine sections, was prepared by a group of authors under the direction of C. W. Horsting. Names of authors are given with the titles of the sections they prepared.

this direction is much poorer than perpendicular to it. As another example: of the three lattices, the B.C.C. structure is associated with the most ductile materials.

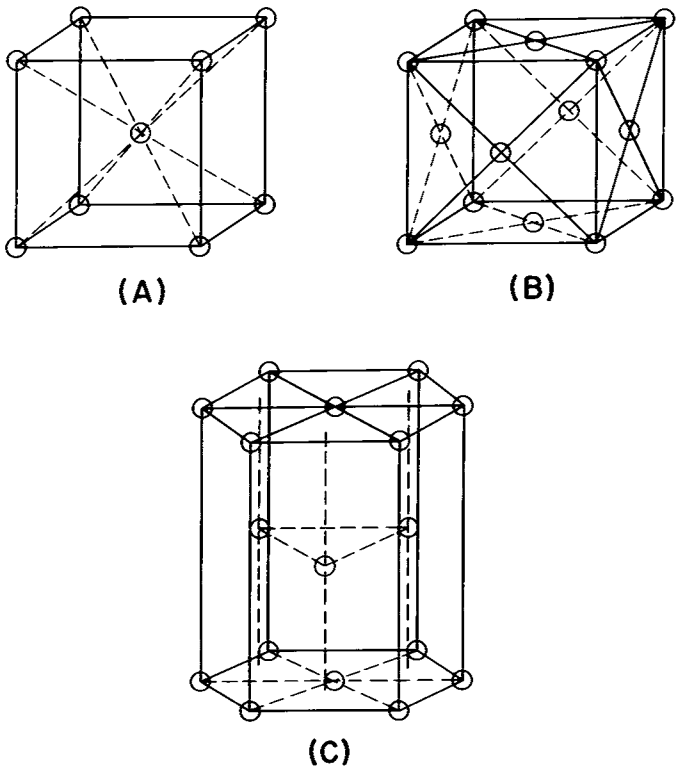


Figure 1. The Three Most Common Metallic Lattice Structures: (A) body-centered cubic; (B) face-centered cubic; (C) close-packed hexagonal

A list of metals and their crystal structures follows:

Face-Centered Cubic
Silver, Gold, Calcium, Copper, Iridium, Nickel, Lead, Palladium, Platinum, Rhodium, Strontium.

Body-Centered Cubic
Barium, Columbium, Cobalt, Iron, Potassium, Lithium, Molybdenum, Sodium, Rubidium, Tantalum, Vanadium, Tungsten.

Hexagonal Close-Packed
Beryllium, Cadmium, Magnesium, Osmium, Tellurium, Zinc, Zirconium.

The problem of work hardening found a new explanation as proposed by Rosenhain[4] who assumed that the first deformation was easy since the glide planes were perfect, but that after some deformation a layer of amorphous material would form at the shear surface. This material would act like sand in a bearing in that it would prevent movement at the glide planes and, therefore, destroy the initial ductility.

GRAIN STRUCTURE

It was Rosenhain, too, who was one of the first to explain the origin of grains. Most metallic objects consist of a large number of small crystals which are united into one solid mass, but are randomly oriented and outwardly show no crystal appearance.

Rosenhain ascribed the origin of this multiplicity of small crystals to the specific behavior of metals during their solidification from the molten state. He assumed that during cooling from the melt a large number of nuclei develop in the cooling liquid metal, that they grow into crystals simultaneously, and that the relationship between the rate of formation of nuclei and their growth into crystals is generally such that a very large number of growing crystals compete for the liquid from which they derive their atomic building blocks.

Since the total number of crystal nuclei is large compared to the total available molten metal, each individual nucleus can grow only into a small "grain" before the melt has solidified completely. The word "grain" is used here instead of crystal because the interference between the external surfaces of the individual crystals which grow from the liquid prevents the development of regular crystal faces. Since the grain orientation is generally random, each grain is surrounded by a zone of disorientation resulting from the misfit of the individual unit crystals as they meet during their growth. This zone of disorder is called the "grain boundary." It manifests itself, among other things, by a preferred tendency to dissolve in chemical etching agents and thus to produce the well-known grain structure in a polished and etched metallographic specimen. A picture of a section of a piece of nickel, polished and etched to show grain boundaries, is presented in Fig. 2.

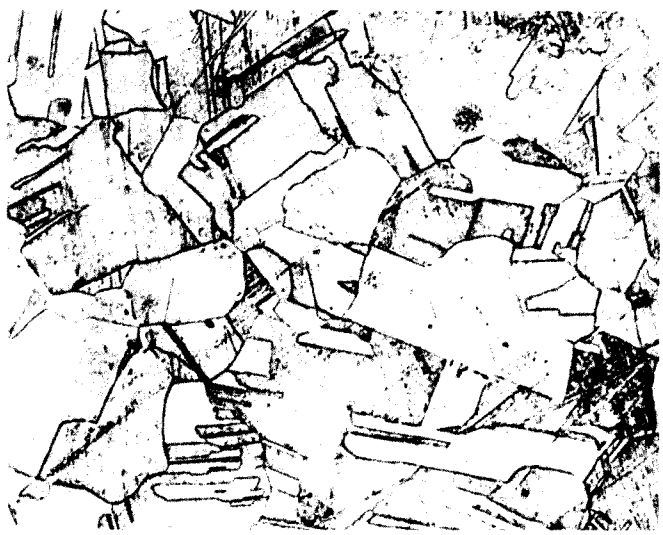


Figure 2. Grain Structure of Nickel (200 X)

In the metal specimen shown, there are approximately 500 grains per square millimeter of cross section. In steel a "grain count" of 10,000 grains per square millimeter is normal. The individual grains in that case are about 0.01 millimeter, or 4×10^{-4} inches, in diameter. Because the interatomic distances are of the order of 2×10^{-8} inches, it follows that there are

2×10^4 atoms lined up across a grain or 50 to 100 million per linear inch. These figures give a good idea of the very small size of individual building stones of which a metal is composed.

Since the rate of nucleus formation and of crystal growth are both temperature-dependent, the number of grains per unit volume can vary over a wide range. The actual size of the grains which result from solidification of a molten metal depends upon the temperature at which the ingot is cast, the cooling rate, and the nature of the metal. Generally speaking, rapid cooling leads to a fine grain structure. In actual practice, the grain can be modified by "working" of the metal, and subsequently by heat treating. Therefore, the control of grain size in a metal is not dependent only on the melting and cooling procedure. This is fortunate because grain size is a very important practical factor in every-day metal technology; it has a direct bearing on the strength of a metal and on the ability of the metal to be formed or shaped. For instance, if strip material is to be used for deep drawing, as it is in the manufacture of kinescope cathodes, it is important to establish a maximum grain-size limit. Small-grained material generally exhibits a uniform deformation pattern because the grain size is chosen such that it is small with respect to the material thickness. If grain size is allowed to approach or equal the thickness of the metal to be deformed, irregularities will occur because the preferred slip direction of the individual grain may cause unwanted local deformations. Surface roughness in deep-drawn parts, often called "orange peel" effect, can be caused by excessive grain size.

Coming back to the atomic picture, through X-ray analysis it was known that the interatomic distances in the space lattice are of the order of 2×10^{-8} inches. The exact distance between neighboring atoms in the space lattice is a function of temperature and pressure and is characteristic for each element.

ALLOTROPY

X-ray analysis also showed that some metals can exist in more than one crystal structure. Below a certain temperature they have one lattice; above it, another. Such metals are called "allotropic." Iron is an example of an allotropic metal. It is body-centered cubic below 910 C and face-centered cubic above this temperature.

The practical importance of allotropy can be realized if one remembers that the change from one lattice to another entails abrupt changes in most physical properties of the metal. These properties include specific volume, thermal and electrical conductivity, corrosion resistance, and others. Fig. 3 shows the effect of allotropic change of lattice on the thermal expansion of a manganese-iron alloy containing 5 per cent manganese. The allotropic change occurs at 785 C; its effect is visible as a bend in the thermal expansion curve. The bend results from the fact that each lattice has its own individual density and expansion coefficient. The fact that the curve does not retrace itself upon cooling is due to incomplete return of the material to the B.C.C. structure, resulting in a permanent change of dimensions. This phenomenon is the cause of severe warping upon heating and cooling of materials which exhibit a phase transition if the temperature range involved brackets the transition temperature. The curve would retrace itself only if heating and cooling were carried out extremely slowly so that the system would be in kinetic equilibrium at each temperature. Some examples of allotropy are given in Table I.

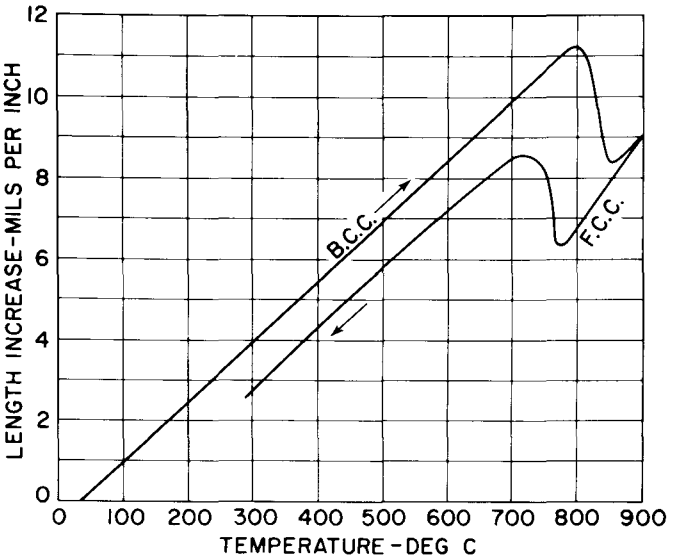


Figure 3. Thermal Expansion Curve of 5 Per Cent Manganese-Iron

Table I

Metals	Low-Temperature Lattice Form	Transition Temperature-deg C	High-Temperature Lattice Form
Chromium	close-packed hexagonal	20	body-centered cubic
Cobalt	close-packed hexagonal	420	face-centered cubic
Iron	body-centered cubic	910	face-centered cubic
Tin	cubic	18	tetragonal
Titanium	close-packed hexagonal	880	body-centered cubic

DUCTILE DEFORMATION

It was probably partly due to the perfect appearance of the X-ray diffraction patterns, that scientists initially went overboard in the assumption of the degree of perfection of the metallic space lattice. But no experimental data or calculations were available to contradict the concept of a perfect crystal lattice.

When, however, a more critical approach was used, several discrepancies between theory and experiment became apparent. For instance, in assuming slip between atom planes as an explanation of the ductile behavior of metals, the mechanism illustrated in Fig. 4 was assumed on the basis of the perfect crystal theory. One should imagine the space lattice illustrated to repeat itself at regular atomic distances in front of and in back of the plane of the paper. Thus planes A, B, C, and D are located parallel to each other and perpendicular to the plane of the paper. When a force is applied as indicated by the arrows, shear is produced between planes B and C. It was assumed that the bonds between the atoms of planes B and C were temporarily disrupted when "glide" took place, and that they were restored after the crystal blocks had moved one atomic distance with respect to each other. This effect was seen as the basic mechanism which allowed the peculiar "ductile" deformation of metals which characterized itself by the fact that the metal remained continuous during the deformation process.

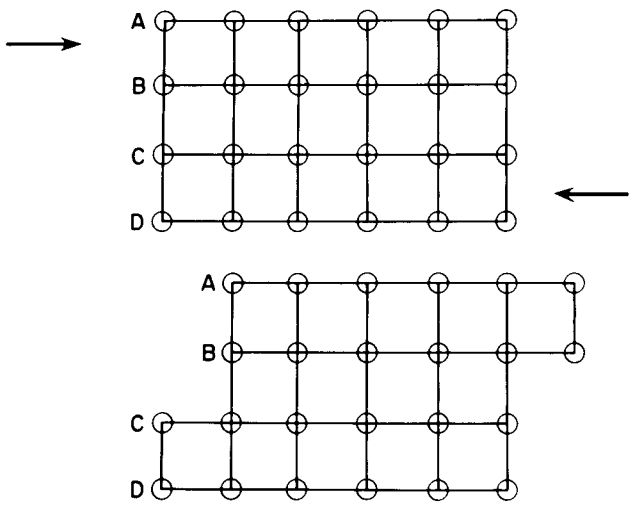


Figure 4. Lattice Slip Mechanism As Originally Assumed

Experimentation showed that the shear stress needed to initiate slip was specific for each metal and that slip took place most easily in planes parallel to the planes which were most densely populated with atoms. Both these factors seemed to fit the theory well and were considered as a confirmation of this line of thinking. When, however, based on the above model, calculations were made of the stress needed to initiate shear, a discrepancy appeared which initially could not be explained. It was found that metals were hundreds of times weaker than they should be if the assumed deformation mechanism was the right one. This finding opened the way to new concepts of the atomic structure of metals. The low strength of the lattice was explained by assuming a less perfect lattice than initially postulated.

DISLOCATIONS

Taylor[5], Orowan[6], and Polanyi[7] almost simultaneously proposed the concept of dislocations to solve the discrepancy between the theoretical and the observed strength of metals.

A dislocation is a lattice defect (there are various kinds) which, basically, allows slip to occur by successive, instead of simultaneous, breaking and restoring of interatomic bonds. In Fig. 5, an attempt is made to explain the operation of slip through a dislocation mechanism. Again, one should think of the lattice plane shown in Fig. 5 as repeating itself in space at regular atomic distances forward and back of the surface of the paper. It is clear that one of the vertical planes perpendicular to the paper is incomplete and that an irregular space is formed in the lattice where this plane ends at atom row a (a row perpendicular to the plane of the paper). This particular type of defect, because it is caused by the edge of an incomplete plane, is called an "edge dislocation." Now, if shear is applied as indicated by the arrows, the bonds between atom rows b and c are broken, and a new bond is established between rows a and c. The result is that the dislocation has moved one step to the left with the incomplete atom plane now ending in row b. This process repeats itself until the dislocation has moved completely across the crystal and a step is formed at the left of the lattice just as in the case of the older model (Fig. 4). Since only a portion of the bonds is broken at one time, however, the energy necessary to initiate shear by this mechanism is much lower than for the older model; calculations of shear forces are no longer in contradiction with observed values. This lack of contradiction was strong support for the dislocation theory; the theory has now grown into a powerful tool which is indispensable in the thinking of today's metallurgists. Many different kinds of dislocation have been postulated and observed; with the aid of the electron microscope, they have actually been seen in action during metal deformation.

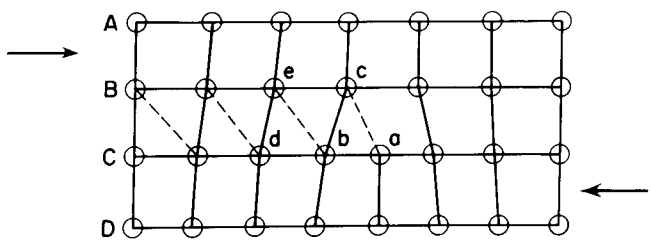


Figure 5. Present Concept of Lattice Slip Mechanism

Whisker growth, an interesting phenomenon that has been discovered in recent years, has further confirmed the concept of lattice imperfections as the source of the low strength of metals. Tiny metallic whiskers were first discovered to grow between the plates of capacitors. They were found to be extremely resilient metallic threads having a tensile strength as high as two million pounds per square inch, a value which, in some instances, was close to the theoretical strength of metals.[8]

Upon investigation, it was found that these whiskers indeed are single crystals almost free of imperfections. Probably sometime in the future man will find ways to utilize the true strength of metals by the use of perfect lattice metal.

TENSILE TESTING

With the background of the atomic structural deformation picture in mind, it is interesting to turn now to the study of deformation on a macro-scale. Since tensile testing is one of the most important everyday means of obtaining information about the mechanical properties of metals, a tensile test curve of a metal will be examined in some detail. This curve shows the change of specimen length as a function of the applied stress when a metal is placed under increasing tension until it fails. The curve of Fig. 6 is that of commercial nickel in wire form.

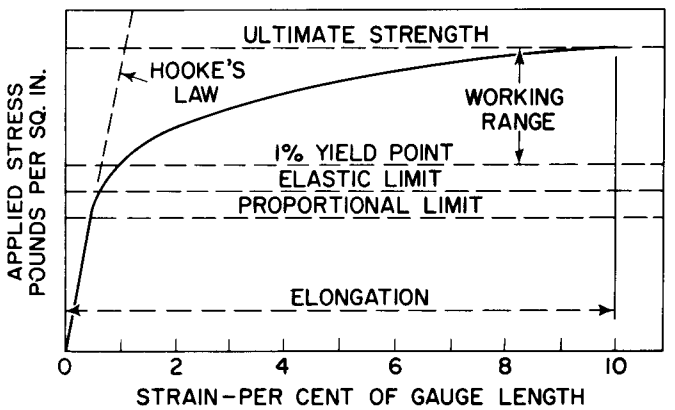


Figure 6. Example of a Tensile Test Curve

It can be seen from Fig. 6 that when force is applied to the specimen, Hooke's law of proportionality is followed in the first part of the curve. In this straight-line portion, deformation is strictly elastic. In other words, if the applied force were released, the specimen would go back to its original length. If the applied force is continuously increased, however, a point is reached which is called the proportional limit (up to that point the atom bonds are only stretched elastically without any of them breaking).

Past the proportional limit, Hooke's law is no longer followed and the tensile curve starts to deviate from the straight line. Very soon after passing the proportional limit, the elastic limit is reached. Beyond this point, deformation starts to become permanent (the dislocation mechanism starts to operate) and, as the specimen is deformed, it elongates and finally breaks.

The maximum stress occurring during the tensile test is called the "ultimate tensile strength." The per cent increase in length of the specimen before breaking as compared to the original length is called "elongation," and is a measure of ductility. Theoretically, the yield point is the same as the elastic limit: the stress at which permanent deformation starts to occur. However, since it is very difficult to determine the elastic limit because the tensile curve often does not show any pronounced abruptness at this point, the yield point has, for practical reasons, been defined as the stress at the 1 per cent strain point, or sometimes at the 0.1 per cent strain point. This definition leaves no doubt about the location of the yield point and is generally sufficiently close to the elastic limit to serve a practical purpose. The distance between the yield point and the ultimate tensile strength is called the "working range." It is the maximum deformation a metal specimen can undergo without failure.

ANNEALING AND RECRYSTALLIZATION

When the micro-picture of deformation was considered, the model showed the dislocation moving through a specimen and finally being expelled at the surface forming a step in the crystal. This picture is basically correct, but it should be realized that it is highly simplified. Dislocations usually move for some distance in the space lattice and are stopped by interaction with obstacles such as impurity atoms and other defects which were not present in the ideal model. This point is important to the explanation of what happens during annealing of a metal. Annealing, as known to practical metallurgists, is the softening of a metal which occurs upon heating and cooling, and which restores its ductility after work hardening. The dislocation theory postulates that, during annealing, dislocations which have become deadlocked redistribute themselves by diffusion into patterns which again allow movement when shear force is applied; in laymen's language, the metal has become softer. New dislocations are probably also formed during the annealing stage.

From the point of view of grain structure, the softening by annealing is accompanied by what is called recrystallization. New grains are formed and, if high temperatures are used, grain growth generally causes the formation of large grains. The crystal-growth part of the structural change does not produce any further softening. Since there is a relationship between the ease of atom movements needed for recovery and the strength of the bonds which tie the atoms together in the space lattice, there is a specific annealing temperature for each metal. The annealing temperature depends on the amount of energy which has accumulated in the lattice as a result of work hardening and upon the presence of impurities or deliberate additions.

Table II lists some metals together with their approximate recrystallization temperatures.

Table II

Metal	Recrystallization Temperature – deg C
Lead	0
Tin	0
Zinc	15
Cadmium	50
Magnesium	150
Aluminum	150
Copper	200
Gold	200
Silver	200
Iron	450
Platinum	450
Nickel	620
Molybdenum	900
Tantalum	1020
Tungsten	1210

DIFFUSION

In the discussion of annealing, redistribution of dislocations, and recrystallization, the idea of atom movements through the space lattice has been tacitly accepted.

When movement of dislocations and other defects in the lattice is described, it should not be forgotten that a defect is actually something negative. It is caused by the absence of atoms from ideal locations. The movement of, for instance, a hole (a lattice position where an atom is missing) in a certain direction is the result of the movement of a series of atoms in the opposite direction: first an atom neighboring the hole moves into the empty position, thus creating a new hole where the atom used to be, and so on. This phenomenon of atom movement through a space lattice is called diffusion; the dislocation theory has been very helpful in explaining many of the phenomena about diffusion which could not be explained before. During the period when the space lattice was seen as too perfect a structure, it was difficult to explain atom movement through a space lattice except by assuming mechanisms like direct exchange or movement between lattice sites (interstitial movement). Other theories assumed simultaneous movements of large numbers of atoms to account for the observed diffusion phenomena. These theories all assumed, however, that there was a balanced flow of matter: when N atoms moved in x-direction, N other atoms moved the opposite way. When the dislocation theory was applied to the diffusion picture, many things that were difficult to explain before became clear. For instance, the activation energy of diffusion was shown to be much smaller than that calculated for the mechanisms of direct exchange or interstitial movement, while the observed activation energy was close to that calculated on the basis of the dislocation mechanism.

Another point that was clearly proven in recent years is that in many cases of mixed lattices (alloys) the diffusion rate of the atoms is specific for each species and independent of the other. This statement does exclude a direct-exchange mechanism. Proof for the independent movement of the individual species was first given by the now famous "marker" experiment carried out by Smigelskas and Kirkendall[9]. They placed inert markers in the form of fine molybdenum wires at the interface of brass and copper (Fig. 7) and, by holding the assembly at high temperature, allowed diffusion to occur. The relative change in position of the markers indicated unequivocally that the zinc atoms from the brass (a copper-zinc alloy) moved faster into the copper plating than the copper atoms from the plating moved into the brass core. Modern radioactive tracer techniques have been very successfully applied to the investigation of diffusion coefficients. Tracer techniques have been especially helpful in establishment of coefficients of self-diffusion, the diffusion, for instance, of silver atoms through a silver lattice. As an example of the wide variety of diffusion speeds, reference is made to Table VII in the section on Cathode Materials (metals) for Receiving Tubes, which contains explicit data on proportionality constant D_0 and activation energies Q of most of the additions that are presently used in cathode nickel.

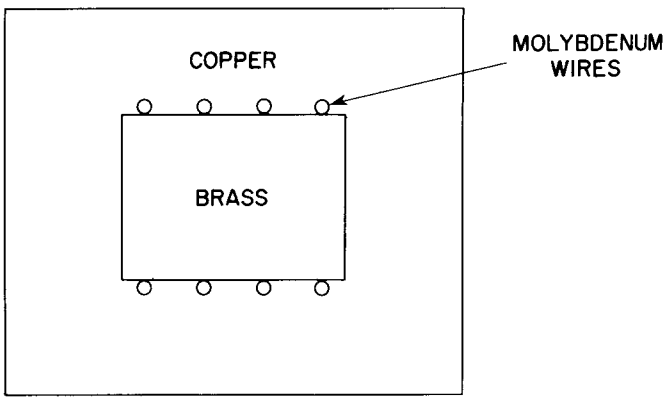


Figure 7. Diffusion Couple With "Markers"

PHASE EQUILIBRIA IN ALLOY SYSTEMS

No discussions of the fundamentals of metallurgy would be complete without mention of phase-equilibrium diagrams of alloy systems. The importance of this subject stems from the wide application which phase-equilibrium data find in today's metallurgical practice. If new alloys are needed which have certain characteristic requirements, data available in phase diagrams are almost invariably consulted. M. Hansen[10] in his book "Aufbau der Zweistofflegierungen" (Constitution of Binary Alloys) has gathered a large volume of such data. Many phase diagrams are also given in the Metals Handbook[11] published by the American Society for Metals. It is not within the province of this article to go into great detail concerning the theory of phase diagrams, but rather to give an over-all discussion of the subject and to refer the reader to the list of textbooks supplied at the end of this section.

For the sake of simplicity, only binary, or two-element, alloys systems will be discussed. A binary phase diagram is a diagram which shows the change with temperature of the structural status of an alloy of two elements for alloy compositions ranging from 100 per cent of the one element, through all possible intermediate mixtures, to 100 per cent of the other element. Each point in such a diagram corresponds to a mixing ratio of the two basic elements and a temperature; the diagram area in which the point is located determines the number of structurally distinct alloy forms present and the chemical composition, physical state, and relative quantities of those alloy forms.

The simplest phase diagram is that for two elements which are completely miscible in all proportions in liquid as well as solid state. The binary system copper-nickel is an example of such a system and Fig. 8 illustrates the case in general. Consider first the extreme ends of the diagram. They represent the phases of the pure elements present at various temperatures. In cooling from the liquid state, the solidification point is passed; at temperatures below this point the liquid becomes solid. If, however, a similar cooling experi-

ment is performed for an intermediate composition, for instance, the one corresponding to the dotted line, no single distinct solidification point is found. In cooling, the alloy passes through an area where liquid and solid are coexistent (colloquially called a mushy range) before it enters the solid area. As the alloy cools through the two-phase area, the amount of solid increases steadily until all liquid has disappeared. The solidification occurs over a range of temperatures instead of at a single solidification temperature as is the case for the pure elements. It should be noticed that the equilibrium composition of the solid and the liquid in the solid-plus-liquid area is found by drawing a horizontal line through the temperature point under consideration. The intersections of the horizontal line with the boundaries of the liquid-plus-solid area give the composition of the solid and liquid which are in equilibrium at the chosen temperature. For instance, composition a, if held at temperature t for a sufficient time to obtain an equilibrium condition, consists of two phases: a liquid of composition b and a solid of composition c. Their relative amounts are given by the ratio of the lengths of the intercepts p and q in such a way that (amount of solid of composition c): (amount of liquid of composition b) = p:q. This relation is called the lever-arm principle. If horizontals are drawn at various decreasing temperatures and the locations of the intersections with the liquidus and solidus curves are observed, it becomes obvious that: (1) the alloy compositions of the solid and of the liquid which are in equilibrium change continuously with decreasing temperature; and (2) that the amount of solid steadily increases at the expense of the liquid. The data needed to construct such diagrams are obtained by determination of cooling curves, by microscopic structure analysis, and by X-ray analysis.

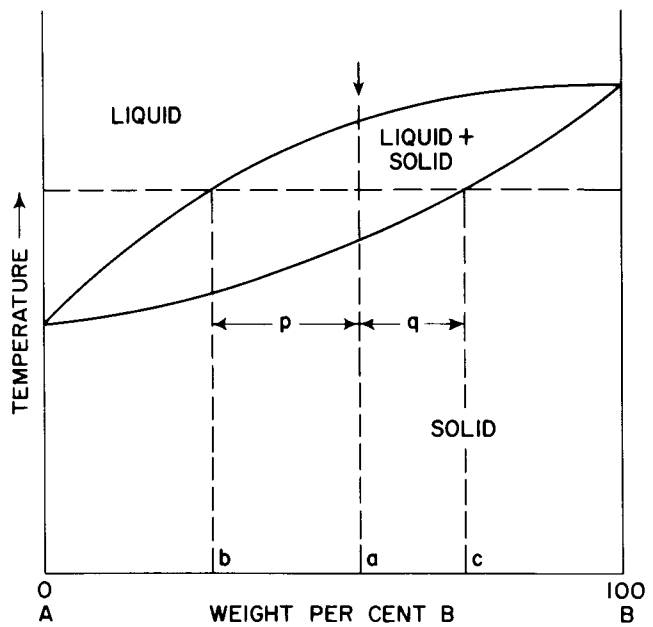


Figure 8. Phase Diagram of Two Mutually Soluble Elements

In the above case of complete solubility of the alloying elements, the solid resulting from solidification of the homogeneous liquid is, in turn, homogeneous and appears under the microscope as a single, continuous type of grain structure. The case of complete immiscibility of the alloying elements in the liquid state as well as in the solid state is the opposite of the one just discussed. It is known to occur in the system molybdenum-copper. A diagram of this type is given in Fig. 9.

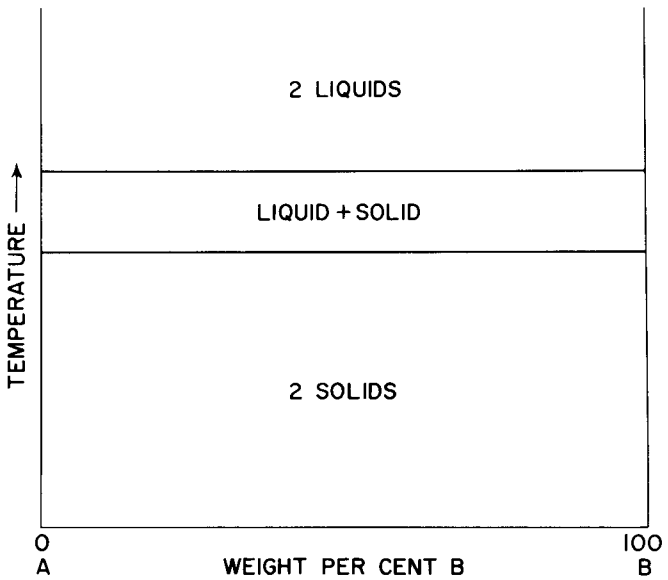


Figure 9. Phase Diagram of Two Mutually Insoluble Elements

The third case is the one where the elements are miscible in the liquid state but not in the solid, as shown in Fig. 10. This type of system is exemplified by the lead-silver alloys. Composition p (Fig. 10) is called the "eutectic" composition. It is the only alloy composition for this alloy system which, like the pure elements, has no mushy range, but a sharp melting point.

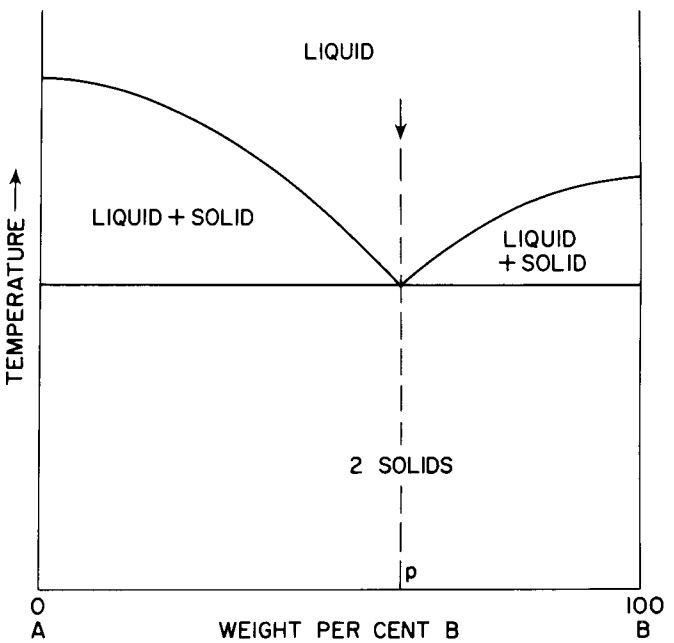


Figure 10. Phase Diagram Showing Mutual Solubility in the Liquid, But Not in the Solid Phase

Microscopic observation of the solid shows two different grain types present in amounts which vary with the alloy composition. The two mixed phases are pure A and pure B.

The phase diagrams of most binary alloy systems are generally combinations of these extreme diagrams. The common type of diagram shown in Fig. 11 features complete liquid solubility and partial solid solubility. The solid shows a mixture of two crystal phases in the two-phase area; however, these are not pure A and B as in the preceding example, but α and β solid solutions. In the solid solution areas, a single homogeneous crystal phase is found. The system silver-copper is an example of a binary system of the type illustrated in Fig. 11.

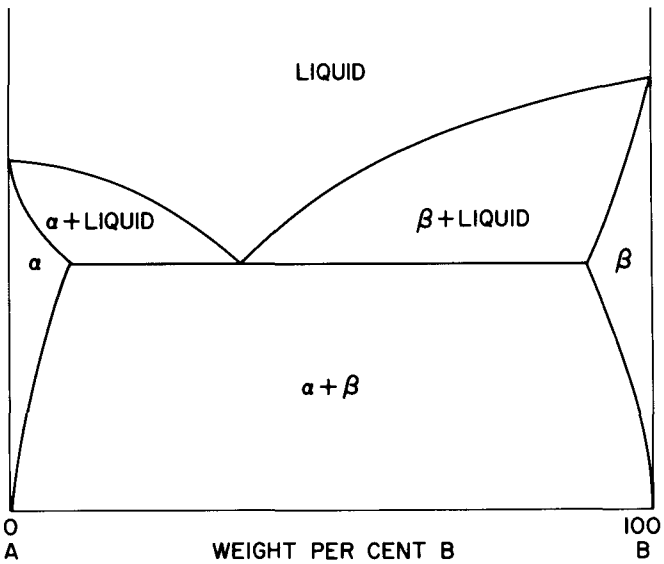


Figure 11. Phase Diagram Showing Complete Liquid- and Partial Solid- Solubility

PHYSICAL PROPERTIES OF ALLOYS

A series of general rules govern the behavior of metals when they are alloyed with others. The chemical and physical properties of the individual metals are generally modified. Consultation of the phase diagram, however, can often throw light on what to expect of the alloy's properties if those of the single elements are known. Fig. 12 gives the general rules concerning the relationship between phase diagram and hardness, corrosion resistance, electrical resistivity, thermal conductivity and lattice parameter.

HARDENING TREATMENTS

Metals can be subjected to various treatments which will increase their hardness. The simplest is work hardening, which is a consequence of ductile deformation. This type of hardness is often only an unwanted by-product of a necessary shaping operation. Other hardening treatments, however, are used just for the purpose of producing high hardness in metals used for springs, files, gears, and many other parts which need hardness to perform their function.

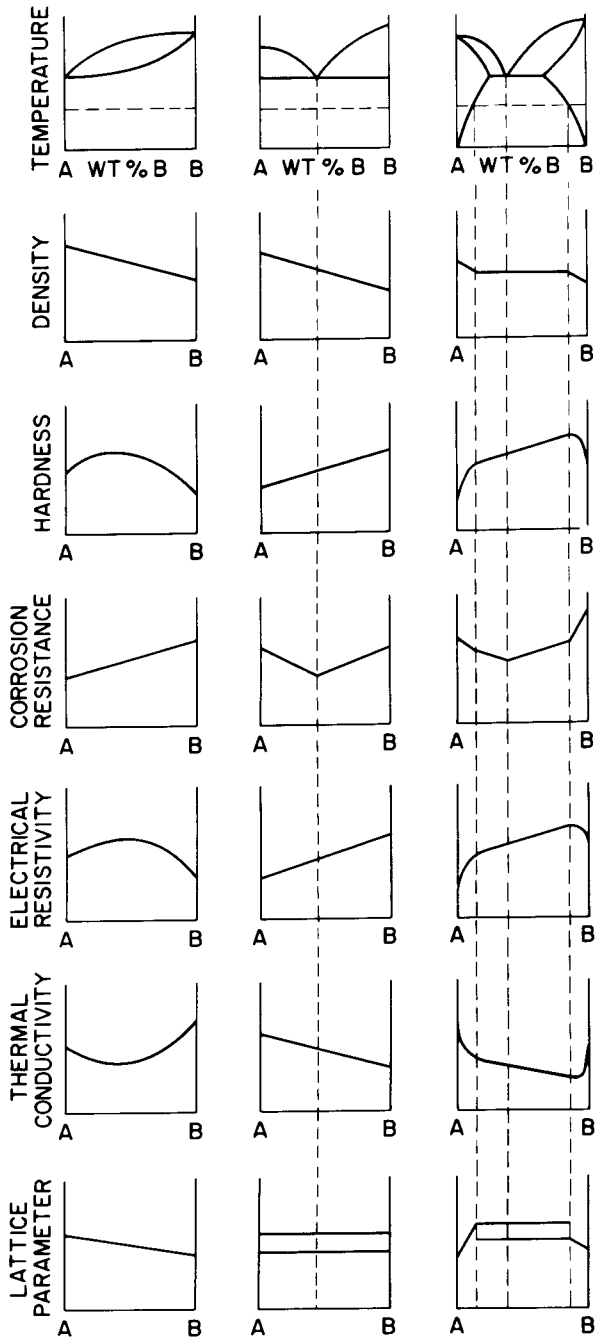


Figure 12. Relationship Between Physical Properties and the Phase Diagram

Two of these treatments are quench hardening and precipitation hardening; both can be illustrated by reference to a phase diagram.

In Fig. 13, the left lower corner of the iron-carbon phase diagram is shown. Two solid-solution fields of carbon in iron are visible at the left.

Ferrite is the solid solution of carbon in α-iron (body-centered cubic); Austenite is the solid solution of carbon in γ-iron (face-centered cubic). Note the allotropic phase transformation on the pure-iron side of the diagram at 910 C. All other areas are two-phase

areas as indicated. The hardenability of steel is due to the fact that the solid solution of carbon in γ -iron is very hard, whereas the two-phase areas are relatively soft.

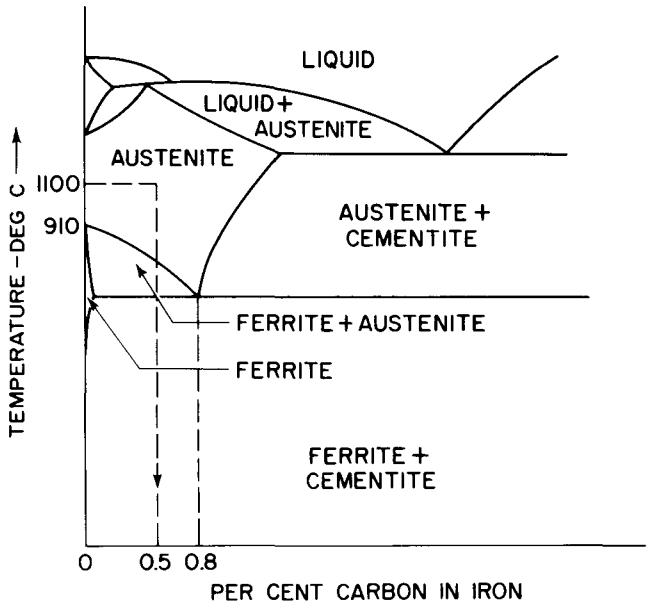


Figure 13. The Iron-Rich Corner of the Iron-Carbon Phase Diagram

Hardening of a steel containing 0.5 per cent carbon, for instance, is done by heating it to 1100 C (see dotted line with arrow); holding it until equilibrium is reached, the ferrite and cementite phases of the two-phase area have disappeared, and the single-phase solid solution has formed; and then quenching the metal so rapidly that, upon its return to room temperature, no time for the reverse reaction to occur has been available.

The other case of hardening, precipitation hardening, is caused by a treatment slightly more complicated than quench hardening; the cause of hardness is not formation of a solid solution, but is the precipitation of a finely dispersed hard phase from a supersaturated solid solution. Fig. 14 illustrates precipitation hardening of beryllium-copper.

In this case the solid solution is soft and can be obtained by a treatment similar to that used for obtaining austenite in the case of steel, that is, by heating, holding, and quenching. The precipitate that causes the hardening is produced by heating the metal to a temperature below the transformation point (within the two-phase area). Equilibrium conditions require that the single α -phase disappear and that in addition to the α -phase, the β'phase appear.

In the case of beryllium-copper, a 1.5 per cent beryllium alloy is generally used; the dotted lines indicate the course followed in the hardening treatment. Note that precipitation hardening of this type is only possible if the borderline between the α -solid solution area and the adjacent $\alpha + \beta'$ two-phase area slants to the right as it does in Fig. 14.

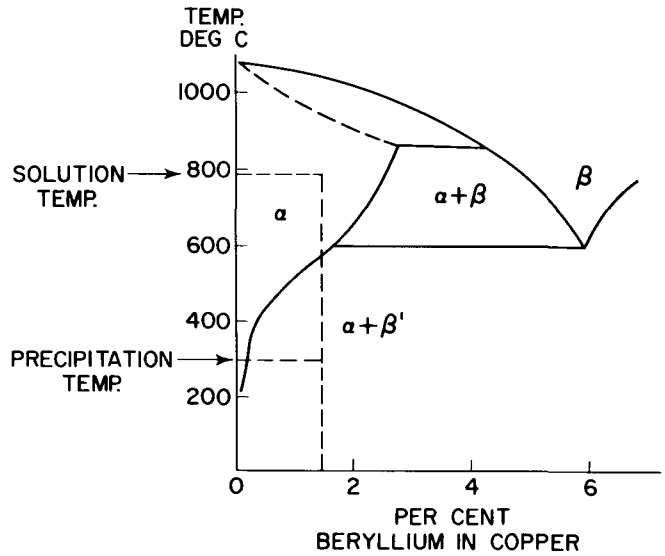


Figure 14. Beryllium Copper, an Example of Precipitation Hardening

VAPOR PRESSURE OF METALS

The vapor pressure of a metal may be defined as: the pressure at which the escape of atoms or molecules from its surface is balanced by the return of these particles from the surrounding gas. Consider, in the case of solids, the random thermal motion of atoms in a crystal lattice. There is a certain probability that an individual atom at the surface may attain the energy necessary to escape the forces binding it to its neighbors. It is known that the higher the temperature, the greater the probability that atoms will escape. In a vacuum, the escaped particles are actually a gas. Some of these particles will strike the surface again and will be recaptured by the attractive forces of the lattice. At any given temperature, the escape of particles is proportional to the surface involved and independent of gas pressure, but the return is dependent on both the surface and the pressure of the surrounding gas. Therefore, the process of escape (evaporation) will continue until it builds up pressure to the point where the rate of condensation is equal to the rate of evaporation. The pressure of gas required to attain this equilibrium is the vapor pressure of the material at the given temperature.

The vapor pressure of any material is largely determined by the strength of the bond which holds atoms within the crystal lattice. Quite often an alloy of a high-vapor-pressure metal low-vapor-pressure metal will exhibit a low vapor pressure because the bonds in the solute lattice are much stronger than those in the lattice of the high-vapor-pressure metal. An example is aluminum-clad iron which is used in electron tubes. The vapor pressure of pure aluminum is quite high, but when it is heated with iron, the rapid formation of an intermetallic compound of high bond strength reduces the vapor pressure of the aluminum to levels which permit its use.

REFERENCES

1. Friedrich, W., P. Knipping, M. von Laue,

Sitzber. math. physikal. Klasse Bayer. Akad. Wiss., Munchen, 1912, 303
2. Bragg, W. H., and W. L. Bragg, The Crystalline State, G. Bell and Sons, Ltd., London, 1933
3. Hume-Rothery, W., G. W. Mabbott, K. M. Channel-Evans, Phil. Trans. Roy. Soc., 1934 (A), 233, 1
4. Rosenhain, W., An Introduction to the Study of Physical Metallurgy, Constable, London, 1915
5. Taylor, G. I., Proc. Roy. Soc. 1934, (A), 145, 362
6. Orowan, F., Z. Physik, 1934, 89, 634
7. Polanyi, M., Z. Physik, 1934, 89, 660
8. Hoffman, G. A., J. of Metals, Vol. 10, No. 9, Sept. 1958, p. 591
9. Smigelskas, A. D. and E. O. Kirkendall, Trans. Am. Inst. Min. Met. Eng., 1947, 171, 130
10. Hansen, M., Aufbau der Zweistofflegierungen, Springer, Berlin (English translation: Constitution of Binary Alloys, McGraw-Hill Book Co., N.Y.)
11. Metals Handbook, Am. Soc. for Metals

TEXTBOOKS FOR FURTHER STUDY

Doan, G. E. and E. M. Mahla, Principles of Physical Metallurgy, McGraw-Hill, New York
Clark, D. S. and W. R. Varney, Physical Metallurgy for Engineers, D. Van Nostrand, New York
Barrett, C. S., Structure of Metals, McGraw-Hill Book Co., New York
Sietz, F., Physics of Metals, McGraw-Hill Inc., New York
Kohl, W. H., Material Technology for Electron Tubes, Reinhold, New York
Hume-Rothery, W., The Structure of Metals and Alloys, Chemical Publishing, New York
Hume-Rothery, W., Electrons, Atoms, Metals and Alloys, Iliffe and Sons, London

METALS FOR ELECTRON TUBES

J. J. Carrona

Manufacture and operation of electron tubes impose a diversity of requirements on the materials used in their construction. Widely varying functions of the metallic members of the tube assembly necessitate individual metallurgical considerations for each case. It is because of these divergent requirements that the tube designer finds some use for almost all metallic elements. Despite these divergent requirements, however, the major uses of metals are restricted to a few elements. In fact, there are only five which comprise the bulk of the metals in use. These metals are nickel, iron, copper, molybdenum, and tungsten. Barium might be listed as the sixth, but its function is not the same as most other metals and, therefore, it is described in a later section on getters. The above five metals, in the pure state, alloyed with each other or with relatively minor amounts of other elements, and plated or clad with small quantities of other metals, make up the electrodes, structural members, conductors, and heaters in electron tubes. These metals are individually discussed in the succeeding pages along with a brief description of some of the less commonly used metals and a few of the more promising new metals.

In a general sense, there are certain requirements which may be described as applying to all metal parts in the electron tube. Foremost among these is the requirement of adaptability to low-cost, mass-production techniques. To be used on a wide scale, the metal must be capable of being formed and welded with conventional fabrication methods. Ductility and good resistance-welding performance are mandatory. A breakdown of the contributory sources of expense in the extremely small parts that are used in tubes reveals that the major portions of costs are those due to fabrication costs rather than to the basic metal cost.

Because parts are designed to conduct electrons and to control electron flow, electrical conductivity is an important property. High conductivity is usually desired, but in some cases low conductivity is needed to provide heating effects.

The environment to which the materials are exposed is one of high vacuum and high temperatures combined with a continual bombardment of electrons. To exist in a stable fashion under such conditions requires that a metal have a low vapor pressure, high hot strength, and reasonably high work function. Values for the properties discussed above may be found in the tables of the last section of this article. Work functions are listed for a few of the elements in Table III (this section). The work functions listed are for the clean element. In the actual cases, barium and barium oxide deposits from the cathode may coat the metal and lower its work function. In fact, if conditions are just right, a good active coating may be deposited and cause the work function to drop to that of the cathode. Certain of the metals (gold, copper, and silver) will combine with the barium and hold the high-work-function values until saturated with barium. As a result, the work function of a particular metal in a particular electron device may range from that of the cathode to that of the clean metal.

NICKEL

Nickel is the most important single metal used in vacuum tubes. In recent years it has been replaced by iron and steel in many applications, but only in the sense that a plated or clad nickel coating on steel is now more common in many tube parts than is pure nickel.

Nickel, in the form of the pure metal, alloys, or surface coating, is used extensively for anodes, grid laterals, grid side rods, inner leads, outer pins, cathodes, filaments, supports, connectors, tabs, shields, glass-sealing leads, and caps.

The mechanical behavior of nickel and high-nickel alloys is as satisfactory from a fabrication sense as is

that of any material. It is ductile and cold-works readily, and it may be drawn or formed into the complex shapes required without special preparation. Its ductility is best evidenced by current factory practice of cold drawing 0.025-inch diameter manganese-nickel wire in a multiple die set-up to 0.002-inch diameter in one continuous operation at a speed of 3000 feet per minute.

Table III
Work Functions of Elements

Element	Work Function—Electron Volts	References
Platinum	5.3-6.3	1, 2, 3
Palladium	4.96-4.99	2, 3
Gold	4.78-4.89	1, 2
Nickel	4.6-5.24	4
Carbon	4.34-4.77	1, 2
Iron	4.23-4.77	4
Silver	4.46-4.47	1, 2, 3
Tungsten	4.44-4.63	4
Copper	4.3-4.45	1, 2, 3
Molybdenum	4.14-4.40	1, 2, 3
Zirconium	3.73-4.50	1, 2, 3, 4
Aluminum	3.0	1
Barium	2.49	4
Lithium	2.21-2.49	1, 2, 3
Potassium	1.9-2.26	1, 2
Rubidium	1.82	1
Cesium	1.67-1.93	1, 2, 4
Oxide-Coated Cathode	1.3	2

Further fabrication ease results from the superior resistance-welding properties of nickel. Although modern welding technology has given adequate means to weld many difficult materials, nickel is still required as an aid in most of the more serious welding problems. Assembly of the mount structure to the stem is still mainly dependent on spot welding, which in most cases requires a nickel surface on at least one of the members to be joined.

Although nickel is not a particularly high-strength material, it possesses enough stiffness to prevent gross sagging or dimensional changes under the temperatures encountered in tube processing or operation. On the other hand, its limited strength restricts its use in the smaller diameters or thicknesses. One seldom encounters a nickel-rich alloy wire or strip of less than 0.001-inch section. As grid-lateral wire-size requirements become smaller and 0.0005 inch wire becomes more commonly used, nickel alloys are being replaced by the refractory metals which can stand up in such service.

The vapor pressure of nickel is low enough to permit heating it to 1000 C without undue vaporization of the metal.[5] Even higher temperatures may be held for short periods with no ill effects. This feature is important because the oxide-coated cathode using nickel as a base must be raised to 1100 to 1200 C during tube exhaust and held at 800 C during operation; plates are raised to 800 or 900 C for outgassing, and grids exposed to cathodes and plates at these temperatures reach similar levels during tube processing.

Chemically, nickel has the advantage of good corrosion resistance which permits storage for long periods with no adverse effects. On exposure to high temperatures in air, the oxide forms relatively slowly. These oxide formations are easily reduced by hydrogen. In miniature stem making, inner leads are oxidized by the fires so that welding of the mount and filaments to these leads is impaired. By placing small inverted cups filled with hydrogen over the oxidized leads while they are still on the stem machine and before the leads can cool, it is feasible to reduce the oxides and improve the welding performance.

Surface-oxide reduction and degassing of nickel are readily accomplished in the same operation. If the part is suitably washed to free the surface of organics and water-soluble compounds, surface oxides may be reduced by hydrogen firing at temperatures of 600 to 800 C. Simultaneously, internally dissolved gases will be replaced by hydrogen. Since hydrogen diffuses rapidly through most metals and particularly so in nickel, it can be quickly removed later by heating the metal in the vacuum of tube exhaust.

The electrical and thermal conductivities of nickel are relatively high. In fact, no other metal, in the same or lower cost ranges, which possesses the strength of nickel, even approaches its conductivity. In leads, plates, grids, and cathodes, this conduction of electrons and heat is important. When good radiating properties are desired, nickel may be blackened or carbonized by passing an oxidized nickel surface through a hydrocarbon gas at a temperature which causes decomposition of the hydrocarbon at the surface of the nickel. In the process, carbon is deposited and bonded. Much plate material is used where a nickel surface has been blackened to provide a radiating efficiency of 80 per cent. Grid wires of carbonized nickel alloys are also available. The increased radiation helps cool the parts and thereby prevents overheating during operation with resulting undesirable emission effects or gas evolution which would seriously shorten tube life.

The most important single application of nickel is as a base for the oxide cathode. In this use, it serves as a support for the structure and a dispenser of the necessary small amounts of reducing agents which activate the coating and maintain cathode emission by producing free barium in the oxide coating. The details of this use are presented later in this article under Cathode Materials (Metals) for Receiving Tubes.

Another rather important application of nickel is for grid lateral wires. A series of high-nickel alloys has been developed through the years which still finds extensive use in electron tubes. Although these alloys are not of the order of strength of the refractory metals, their strength is adequate for many grids. To reduce grid emission effects and/or improve hot strength, the alloy manufacturer adds elements such as manganese, titanium, silicon, aluminum, chromium, cobalt,

iron, molybdenum, and tungsten. Alloys of these types include Magno-nickel, Permanickel, Hastelloy, and RCA N100W.

IRON AND STEEL

Iron and the low-carbon steels most commonly used in the electron tube have properties which are generally somewhat less desirable than those of nickel. Nevertheless these metals are probably used in greater weight in low-cost tubes than is nickel. Obviously this popularity is due to the great cost advantage that these metals hold over all other metals. This advantage persists even though the bare metals are almost never found in use and the relatively high cost of plating or cladding them with nickel or copper has to be added to their costs. Basic metal values along with some costs for finished wire are listed in Table IV for some of the more common metals, and for bare and plated iron and low carbon steel. The relationships given are relative to the cost of SAE 1010 steel and are approximate. They will vary widely both with special processing requirements and with time. They are intended only to illustrate the influence of fabrication difficulty on the cost of small-section metal products. Note that although nickel is more than twice the price of nickel-plated steel wire at 0.040 inch, there is no difference at 0.004 inch. The comparison between basic values and 0.004 inch wire costs for molybdenum and tungsten reflects the poor workability of tungsten.

Table IV
Relative Costs of
Bulk and Fabricated Metals
(as of June, 1960)

Metal	Basic Values	Wire	
		0.040-in. Diameter	0.004-in. Diameter
SAE 1010	1	6	—
Nickel-Plated 1010 Steel	—	22	150
Nickel	18	51	150
Copper	6	27	—
Molybdenum	84	—	660
Tungsten	112	—	1100

The most serious drawback of iron is its poor corrosion resistance. Storage life is extremely poor unless extensive protective measures are taken. It is mainly for this reason that unprotected iron or steel is not used in vacuum tubes. The oxides of iron are objectionable on electrode surfaces and, unless the metal is properly plated or clad to retard oxide formation, it can not be prevented. Electron bombardment of an oxidized iron surface may result in high secondary emission and will promote gas evolution.

Despite the obvious differences between iron and nickel, their physical properties, for the most part, are not too dissimilar. Their vapor pressures are almost identical, the melting point of iron is only 100 C above that of nickel, the thermal and electrical conductivities of iron are 2/3 those of nickel, and strength and ductility are similar. One can see that, except for the poor chemical behavior of iron, it can adequately replace nickel if the lower conductivities of iron are compensated for by some increase in section thickness. The use of iron or steel, with surface coatings to mask their chemical behavior, has replaced nickel in most supports, grid side rods, plates, and inner leads. In most of these cases, nickel is still required on the surface, but with plates or anodes aluminum is often used as the surface metal; such materials are supplying a large share of the market. Copper, as a surface coating on grid side rods and leads, is finding increased use due to the conductivity improvement it contributes.

Gas evolution from iron and steel has been a problem during tube operation. It is known that these metals contain impurities of carbon and oxygen. The latter may be combined as compounds which can be reduced by carbon. At the temperatures of 800 to 900 C reached during exhaust, the carbon may diffuse to and reduce the oxygen compounds. The gas (probably carbon dioxide) is continuously evolved during this period by diffusion to the surface. When the plate is cooled, the gas-producing reaction is stopped as is the loss of gas by diffusion. As a result, the metal is saturated with gas which is released during the life of the tube. There is some question as to whether reaction between oxides and carbon proceeds at significant rates at the operating temperatures of 300-400 C.

The first attempts to eliminate these gases consisted of reducing the carbon content to restrict the reaction. Resultant products of relatively carbon-free iron (Armco or Svea Iron) were partially successful. Carbon is reduced to about 0.03 per cent or less, and subsequent annealing and hydrogen firing will decarburize the iron still further. In thin plate materials, long decarburization firing is capable of almost completely eliminating all carbon with attendant decreases of gases evolved on life. Some tube manufacturers are presently exploring decarburization and have found that the presence of less than 0.01 per cent carbon in an anode gives negligibly low gas evolution on life.[6]

Another method of restricting the gas-producing reaction is to chemically combine the oxygen in the iron or steel in compounds so stable that they cannot be reduced by carbon. This method is presently standard for most of the steels used in the electron tube. During the casting of the ingot, aluminum is added to the ladle of molten steel. The resulting product is aluminum-killed steel. The oxides of aluminum which form are of the desired stability. Unfortunately, those oxides of iron which occur on the surface on subsequent heat treatments are not combined with the aluminum and may present gas problems. At present, many parts are still made of normal low-carbon steels although there is a general movement to convert to aluminum-killed steels.

COPPER

The high electrical and thermal conductivities, softness, and chemical behavior of copper are properties which find applications in vacuum tubes. The use of copper is restricted by its softness and fairly high vapor pressure (see table in last section of this article), but since, among low-cost metals, it is unexcelled as a

conductor, it is widely used in special applications. These uses include glass sealing, grid side rods, leads, plates (as a conductive core), and as an exhaust tubulation and pinch-off.

The high conductivity of copper permits its use in leads and grid side rods where an attempt is made to facilitate heat removal to cool electrodes. Leads are very often made of copper to provide good electrical conductivity for high-frequency currents. The superior solderability of copper is one basis for its extensive use in outer leads of octal-base tubes where the leads are soldered to base pins. Adequate seals are made in certain tubes by mechanical pinch-off of a copper exhaust tubulation. The softness of the metal allows the deformation and cold welding of the tubing to form a vacuum-tight seal in a simple fashion.

Recently, copper has gained some notice as a core for plate materials. Poorly conductive anodes in beam-type tubes run hot in the area of electron-beam impingement because the heat generated in these areas cannot readily flow to cooler sections by conduction. A core of copper representing 40 per cent of the section thickness distributes the heat throughout the plate and sharply reduces the maximum temperatures reached in the areas where heat is generated. This temperature reduction results in a reduction of gas evolution from the plate and extends tube life.

Copper, properly prepared with a thin (0.0001 inch) oxide coating, seals well to soft glass. The solubility of copper oxide in glass, and the good adherence of the oxide to the metal promote good glass-to-metal bonds. Unfortunately, the extremely high thermal expansion of copper does not permit its use for sealing in other than very thin sections where thermal mismatch is minimized or overcome by deformation of the metal. A compensation for the excessive expansion coefficient is made by cladding a low-expansion 42 per cent nickel-iron core with copper. A rather close match to soft glass results and this product, known as "dumet," has been the standard low-cost lead-sealing metal since early in the century.

The restrictions placed on the use of copper by its very low strength may be modified by alloying it with other metals. Small amounts of metallic additions give appreciable increases in strength. Note, in the last section of this article, the 15 per cent increase in tensile strength that results from the addition of 1 per cent chromium. These additions may also cause excessive reduction in conductivity. Although the effects of elements such as chromium, silver, and tin are not too severe, most elements, especially oxygen, have a serious effect on conductivity. It is for this reason that the standard grade of copper for all electrical conductors is the OFHC (oxygen-free-high-conductivity) grade. Oxygen is carefully removed by the addition of elements to the melt which chemically tie up the oxygen. The use of different deoxidizers leads to the various grades of OFHC copper-phosphorus-deoxidized, silicon-deoxidized, and others.

TUNGSTEN

The refractory nature of tungsten is the reason for most of its use. It has the highest melting point of all the metallic elements. Strength at elevated temperatures is phenomenal. Combined with these properties, it is moderately conductive, has a low vapor pressure, low gas content, and low thermal expansion. It is used extensively as heater wire and for very fine grid wires.

Because tungsten cannot be cast successfully, powder metallurgy is used to produce it. A fine tungsten powder having particles ranging from 1/2 to 10 microns in size is compressed into bars under tremendous pressures. These bars are prefired in hydrogen at 1500 C to make them strong enough for handling. After this operation, the bar is placed into a sintering bottle where a current of 3000 amperes is passed through a 1/2-inch-square bar section. The temperature of the bar is raised close to the melting point of tungsten so that the powder sinters into a dense structure. The bar is then strong enough for working and is hot-swaged into a round section. By further swaging, annealing, and hot drawing, the rod is converted to wire. Wire as fine as 0.0005 inch diameter is drawn; smaller sizes are chemically etched to, at present, a commercial minimum of 0.00025 inch diameter. The tensile strength increases as the size is reduced and reaches 500,000 pounds per square inch—a value considerably higher than that of any other pure metal. This high value combined with the highest modulus of elasticity, makes tungsten the stiffest and strongest material available.

The structure of tungsten wire is a fibrous, fine-grained one. If the temperature of the metal is raised above 1500 C, recrystallization occurs and large grains form rapidly. These large grains, which may approach the wire size, have very weak and brittle boundaries. This recrystallization causes a sharp drop in strength and causes many of our heater failures. Early in the history of the manufacture of tungsten, the good emission properties of tungsten doped with 1 or 2 per cent of thorium oxide were noted and made use of in power tubes. It was found that the thoria retarded catastrophic grain growth. After much experimentation, other doping agents were found. Present practice is to use very small amounts of calcium, aluminum, and silicon as the grain refiners in the manufacture of non-sag tungsten. Because these elements vaporize at high temperature stages in the production process, and the amounts remaining in finished wire are somewhat variable, the occurrence of brittle heater failure due to recrystallization is, even now, frequent enough to cause concern. Alloys with rhenium are being developed as a substitute not exhibiting this fault.

Developments in grid wires have reached the stage where the most advanced tube designs are almost wholly with structures using laterals of 0.0003 inch to 0.0005 inch diameter brazed to the equivalent of side rods. Frame grids and small, cylindrical multi-side-rod grids similar to those used in pencil tubes will become more common. The most satisfactory metal for laterals for these grids is tungsten. It is probable that no other metal, except perhaps molybdenum, can withstand the extremes of stress and temperature that these structures must undergo in fabrication and tube operation.

Tungsten has the lowest thermal expansion coeffi-

cient of any metal—about 4.6×10^{-6} inches per inch per degree centigrade. It closely matches the expansion of hard glass and, with its excellent conductivity and reasonable oxide adherence, makes a good lead-sealing material for hard glass. Due to its relatively high cost, however, it is used as lead-sealing material only in higher-priced tubes.

There are several disadvantages to the use of tungsten. It has little ductility and cannot be cold worked to form complex shapes. Normally, such a restriction would void its use in low-cost tubes, but its high-temperature properties are required even at the expense of this deficiency. The use of tungsten is limited to bend shapes since the best elongation obtainable is two per cent. Fortunately some relief of the situation may be expected; experimental quantities of pure tungsten with a ten per cent elongation have been produced.

Another difficulty with tungsten is the embrittlement that the metal undergoes when contaminated with small amounts of nickel. When tungsten is welded to nickel in areas which subsequently become heated, diffusion will introduce nickel to the tungsten and produce failures adjacent to the weld. As a general rule, intimate contact of tungsten with nickel (or iron, or cobalt) at elevated temperatures should be avoided to prevent such failures.

MOLYBDENUM

The properties of tungsten, if reduced in value somewhat, would adequately describe those of molybdenum. The major differences lie in the appreciably better ductility and lower strength exhibited by molybdenum which permit it to be cold drawn. Only a few years ago, most molybdenum wire was produced by hot drawing, but now much of it is made by conventional cold drawing. It is manufactured by the same powder methods used for tungsten, but in recent years more and more is arc-cast by the consumable electrode process. This process consists of maintaining an arc between an electrode of molybdenum and a water-cooled metal crucible. The heat generated by the arc causes the electrode to progressively melt and fall into the crucible. The cold crucible freezes the molten molybdenum preventing attack on the crucible metal. An ingot so cast is reduced in the normal fashion.

Most molybdenum in electron tubes is used in grid wires, supports, and high-temperature elements of complex shape. It performs admirably in service as a grid lateral wire since it has good thermal conductivity; it has good strength and is ductile enough to permit sizing and forming of the grid; it is readily cleaned because its oxide is volatile, and the metal itself has a low vapor pressure.

The volatility of molybdenum oxide is both an advantage and a disadvantage. It is a simple matter to clean a molybdenum grid in a vacuum because all that is necessary is to heat it to about 600 C at which temperature the oxide vaporizes. At the same time, this oxide because of its injurious effect on emission, should be kept away from the cathode. The problem of removing the oxide and keeping it from the cathode is usually solved by keeping the cathode hot so that no oxide will condense on it.

Because of its freely oxidizing behavior, molybdenum grids are usually gold plated. Barium and its compounds (from the cathode) are insoluble in molybdenum, and deposit on the grid during tube operation. Reduction of the compounds by the molybdenum oxide produces free barium with resulting grid emission. A small amount of gold restricts the formation of oxides and, because barium dissolves readily in gold, the problem of grid emission is reduced. Where a gold-plated surface is desired in a grid, only molybdenum or tungsten can be used as the base metal. Gold shows appreciable solubility in most metals but practically none in molybdenum or tungsten. During the period that high temperatures are experienced by the grid in processing and in operation, gold will diffuse into the base metal and lose its effectiveness if nickel or a similar metal is. used, but will remain unaffected on molybdenum or tungsten. Figs. 15, 16, and 17 show this effect.

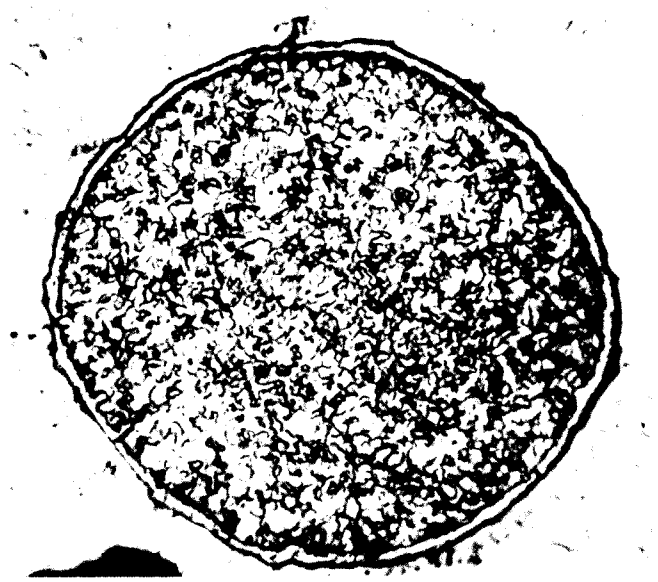


Figure 15. 0.0033-Inch Diameter Gold-Plated Molybdenum Wire After 500 Hours as a Grid Lateral: the gold plate is still concentrated at the surface

A major use for molybdenum is as a disposable mandrel on which tungsten heater coils are wound and set by firing. The molybdenum mandrel can be removed by etching, leaving the tungsten coil. No other metal of lower or equal cost can withstand the setting temperature without contaminating the tungsten.

Molybdenum is often alloyed with tungsten in an attempt to impart some degree of the ductility of molybdenum without sacrificing too much of the high strength of tungsten. Some grid wires in current use are made of a 50 per cent tungsten alloy (Dowmo D18). Tungsten is sometimes used to strengthen molybdenum. Twenty and twenty-five per cent (H3, M8) tungsten alloys have seen some service in grids. Since the addition of molybdenum increases the tendency for tungsten heaters to become brittle on operation, use of such alloys has largely been discontinued for heater service.

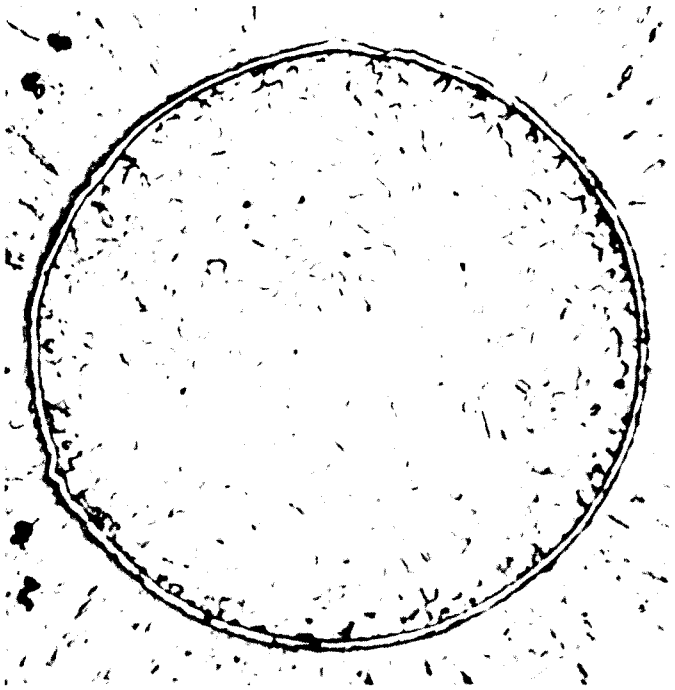


Figure 16. 0.0033-Inch Diameter Gold-Plated Manganese-Nickel-Alloy Wire Before Tube Processing: the gold plate is the thin outer layer shown

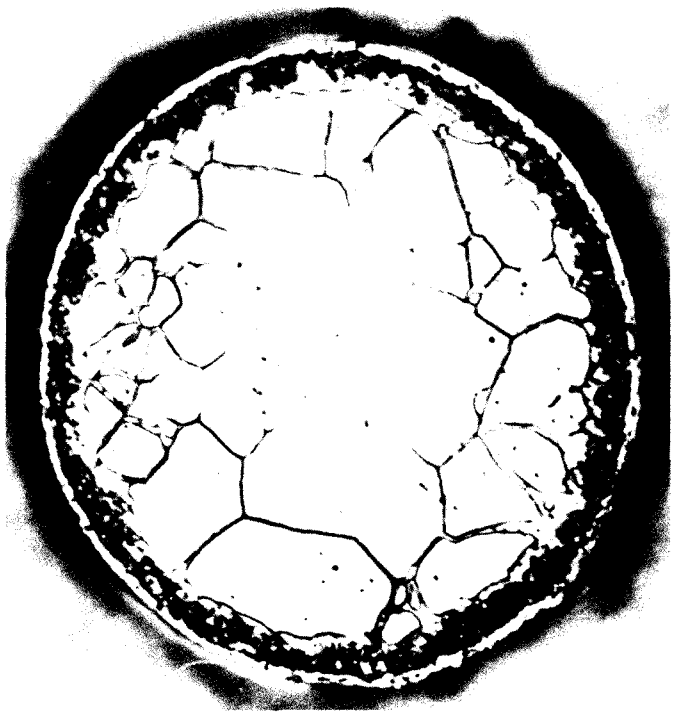


Figure 17. Wire of Fig. 16 After 500 Hours of Tube Operation: the gold has diffused into the base, especially along the grain boundaries; at no place in this section is there enough gold concentrated to exhibit the characteristic gold color

OTHER METALS

Silver

Silver approximates the properties of copper, but is considerably more expensive and undergoes a different type of corrosion. Because of its poor structural strength, it is seldom used except as a surface coating on stronger base materials. Because it has excellent conductivity and does not form high-resistance corrosion products, it is used as a surface coating on parts which must conduct the skin currents associated with high-frequency signals. A combination of the high-frequency use and use for its glass-sealing quality is found in pencil tubes where silver-plated steel parts are sealed to glass for tube assembly and are operated at a very high radio frequency. The high solubility of barium in silver and the insolubility of silver in nickel promotes the use of silver as a coating on nickel-alloy grid wires where the silver will not diffuse into the base and lose its effectiveness. The vapor pressure of silver is somewhat higher than that of gold, but at 1/5 the cost of gold, silver finds considerable use on nickel-alloy grid wires.

Aluminum

Most aluminum in tubes is used in plates and getters. Aluminum-clad steel finds extensive use as a plate material. When a thin layer of aluminum on iron is heated, an iron-aluminum compound forms. During the reaction, a dark, rough surface with an emissivity of 70 to 80 per cent forms. The aluminum-iron structure resulting has bonds of adequate strength to prevent undue vaporization of the normally volatile aluminum.

Getters of aluminum-clad barium are common. The aluminum protects the barium during storage and reacts exothermally to form the getter flash. Further uses are restricted by the low melting point and fairly high vapor pressure of the pure metal. It is found in tubes only when combined with other metals.

Chromium

Pure chromium finds no use in electron tubes but is commonly used in alloys for three reasons. Chromium is added to glass-sealing alloys of the 45 per cent nickel, 49 per cent iron, 6 per cent chromium types (RCA I-63 Number 4, etc.) to form oxidized surfaces which seal readily to glass. Chromium greatly increases the strength of alloys of nickel or iron. The splendid oxidation resistance imparted to alloys containing chromium is of interest. The thin protective film of oxides of chromium which forms spontaneously gives the corrosion resistance of the stainless steels and Nichromes.

The high vapor pressure of chromium would prevent its use even in alloys if it were not for the fact that the oxide layer which may be found on its surface suppresses the vaporization of the metal.

Tantalum

Where a melting point and vapor pressure approaching those of tungsten are desired, with sufficient work-

ability to permit drawing, tantalum may be used. It is expensive and must be handled in extremely dry inert gases of vacuum for high-temperature processing and, accordingly, is used somewhat sparingly.

Titanium and Zirconium

Zirconium has long been used as a continuous getter in power tubes. Some use has been made of its properties of reaction and combination with all but inert gases at temperatures above 800 C in the receiving tube. If kept at such a temperature, it can react with, and absorb by diffusion, the gases which evolve on tube life. Titanium has received considerable emphasis in aircraft structures and, due to the increased use and government subsidies, has seen a radical price drop. Since most of its properties are extremely close to those of zirconium, it has been used to replace the latter with considerable economy in many instances.

REFERENCES

1. Kohl and Rice, "Electron Tubes for Critical Environments," WADC Technical Report, March 1958, PB 131852 Office of Technical Service, Dept. of Commerce, Washington 25, D.C.
2. Hermann and Wagener, Oxide-Coated Cathode, Chapman & Hall, 1951
3. Dow, Fundamentals of Engineering Electronics, Wiley, 1952
4. Smith, "Thermionic Emission" (Chapter 6), Handbook of Physics, McGraw-Hill, 1958
5. Jackson and Jenkins, "Nickel in Electronics," Metallurgia, Manchr, June 1953
6. Kerstetter, ASTM Symposium on Cleaning of Electron Device Components and Materials, October 1958
7. Miller and Millis, "The Production of Aliron Clad Copper and the Application Advantages in Receiving Tubes," 4th National Tube Techniques Conference, 1958
8. Metals Handbook, American Society for Metals, 1948
9. Smithells, Metals Reference Book, Interscience, 1949

TESTING OF METALS

C. H. Li

Metals for use in electron tubes are tested by various methods to ascertain their qualities or properties. These test methods include:

Mechanical testing
Metallographic testing
X-ray testing
Thermal testing
Chemical testing
Electrical testing
Miscellaneous testing

MECHANICAL TESTING

The following mechanical tests are standard production tools for evaluating the physical properties of metals before, during, or after processing. They detect inferior materials and enable their rejection prior to their being processed or built into tubes. These tests, therefore, not only lower the cost of tube manufacture, but also improve tube quality.

Tensile Tests

Metal wire and strip materials are tested on various testing machines for tensile properties. A testing machine is a device for producing a desired distortion in the specimen to be investigated. The testing machine has a means of holding the specimen; a means of producing, at a proper speed, a distortion of the kind desired; and a reliable device for measuring the resistance offered by the specimen. The following items relating to tensile tests are important in tube applications:

1. A stress-strain diagram is a diagram plotted with values of stress (or applied force) as the ordinates and of strain (or resulting deformation) as the abscissas.

2. The yield strength is rigorously defined as the stress at the point on the stress-strain curve where the curve departs from linearity. At this point, deformation ceases to be elastic. Because it is usually impractical, and probably impossible, to determine the stress at which inelastic action begins, it is customary to define the yield strength as the strength at a definite value of strain – usually 0.1 per cent.

3. The tensile strength is the maximum tensile stress which the specimen can withstand.

4. The elongation is the maximum tensile strain which the specimen can stand without breaking. It is usually expressed in percentage of the initial specimen length.

5. The working range is the difference between the yield and tensile strengths. It is very important in determining the performance of grid wires on grid winding and the quality of the resultant grids, since this is the range of deformation available for the grid-making processes such as wire stretching and grid sizing.

Simple special devices have been designed and built for measuring either elongations or tensile strengths only. They are not accurate, but are very convenient for factory use.

Among the metals tested for tensile properties are: grid and heater wires, grid side rods, cathode strips, and various semifinished products such as nickel-plated iron wires.

Hardness Tests

A hardness test consists mainly of applying a specified load to the surface of the testing specimen through

a standard indentor, and measuring the size of the resulting permanent impression. Several hardness tests are in use in the tube industry. The Brinell, Rockwell and Superficial Rockwell hardness tests are employed for cathode nickel, anode plate materials, and other strip metals. The Vickers hardness test, which is more accurate but slower than the other tests, and the Tukon micro-hardness method are mainly used for calibration or research work.

Stiffness Test

Metal wires and strips are often tested for stiffness or hardness by being bent as cantilevers under specified bending moments. The angle of bend decreases with increase in the stiffness or hardness. This method of testing has wide applications including the control of annealing, drawing, and working of metal wires and strips. An instrument built by RCA (present model No. 719H) is being used widely in the industry.

Springback Test

The springback test measures the angle of springback of a metal strip after being released from a 90 degree bend across the grain (perpendicular to the direction of rolling). Cathode strips are being so tested to eliminate soft strips that are difficult to feed into cathode-forming machines, and hard strips that can not be readily formed without cracking.

Ductility Test

Cathode and plate metal strips are tested for ductility, or deep-drawing properties. In one form of the test, a disc of the strip is rigidly clamped between two plates. A spherical punch is forced into the disc until the disc fails by rupture. This test subjects the disc to radial tension. Ductility is measured by the depth of penetration of the spherical penetrator at the time of rupture.

METALLOGRAPHIC TESTING

The primary object of metallographic examination is to reveal the constituents and structures of metals and their alloys by means of the microscope. Because of the limited area of examination, the proper selection and preparation of specimens are of major importance.

Specimens too small to be handled readily during polishing are mounted to insure a surface satisfactory for microscopic study. The specimens may be mounted by mechanical means, by electroplating, or by embedding them in thermosetting or thermoplastic materials.

Specimens are preferably 1/2 to 1 inch in diameter for easy handling. Their preparation involves several steps:

1. Cutting a properly selected specimen without affecting its structures.

2. Cleaning and grinding one surface of the specimen on progressively finer emery papers.

3. Polishing the ground surface with very fine polishing media, or electrolytically.

4. Etching the polished surface chemically or electrolytically to bring out the desired structure.

Special precautions for grinding or polishing different metals and etching agents are given in the ASM Metals Handbook. *

Other important applications of metallographic study are the determination of: (1) inclusions in cathode nickel alloys; (2) surface oxides; and (3) impurities in clad and electroplated parts. Often, chemical determinations of impurities would not be adequate, since they would not reveal their type, size and distribution.

Another important application of metallographic study is the determination of the grain size of metal specimens during various stages of processing to insure the desired finished products.

X-RAY TESTING

X-ray testing usually requires only minute samples. Such tests are often non-destructive, which may be a distinct advantage.

Industrial Radiography

X-rays are used to study structures by showing shadow pictures of metals more or less transparent to the radiation. They are thus employed to compare competitive tube assemblies or parts, and to locate hot, cold, or fatigue cracks; and inclusions, segregations, or other heterogeneities. Gas holes, porosities, or microshrinkage in castings, in weldments, and in metal parts can also be detected.

Diffraction Study

X-rays are scattered, or diffracted, from regularly arranged atoms of metal crystals. The diffracted rays are recorded on photographs for various purposes, such as:

1. Identification of metals such as tungsten or nickel

2. Determination of crystal structures, orientation of crystals, and the space lattices and space groups

3. Measurement of stress and strain, e.g., in tungsten heater wires

4. Study of phase transformations and equilibrium relations

5. Chemical identification of thin surface film, growth and orientation of crystals in such film

6. Study of fatigue, cold work or machining, electroplatings and age hardening

7. Study of heat treatments, recrystallization, and

*The American Society for Metals, 1948

grain size (e. g., on Nipron or nickel-plated iron grid wires)

THERMAL TESTING

Dilatometric Analysis

The purposes of dilatometric analyses are:

1. To find the coefficients of thermal expansion, and thus, to determine whether a metal "fits" a glass, a ceramic, or another metal to which it is to be joined

2. To find critical temperatures at which changes in metal structures occur

3. To make simulated service tests on electron tube parts or assemblies

4. To check metal compositions, for example, to distinguish between iron and nickel-iron alloys

The analysis consists of determining the thermal expansion and contraction by measuring the length and temperature of the specimen. The rate of change of temperature is often important, in which case a measure of time is necessary. A program-temperature controller is used to insure a maximum change of 83 C per hour within the transformation range. To guard against chemical changes of specimens during test, argon or helium gases are often used. Vacuum impedes heat flow and its use might, therefore, result in erroneous temperature readings.

Emissivity Measurements

Heat radiation properties of plate, grid, cathode, and heater materials are often a critical factor in tube design. One way of measuring these properties is to heat a sample of standard shape and size, and determine the cooling rate in vacuum at different temperatures. An automatic recorder is convenient for this work; the recorder is connected to one end of a thermocouple which has its other end welded onto the sample.

Thermal Analysis

Thermal analysis is the process of determining the temperature at which changes in atomic arrangement accompanied by a usually abrupt change in heat content occurs. Typical examples of these changes are the fusion and solidification of metals, and phase or magnetic transformations. Melting points of different brazing or soldering alloys and Curie temperatures of magnetic alloys may thus be determined. Essentially, the analysis consists of closely following the temperature of the specimen as it is heated or cooled with a constant heat input, and plotting the cooling or temperature-time curve. The abrupt changes in heat content at the critical temperature are generally observable as breaks in the temperature-time curves.

CHEMICAL TESTING

Compositions Analysis by Chemical and Spectrographic Means

Qualitative and quantitative analyses of various metals are covered in appropriate ASTM (American Society for Testing Materials) standards.

Rapid Identification of Metals

Metals are often identified by the color of a properly prepared solution, presence or absence of residues, or precipitation with chemical reagents. These methods have been used to distinguish nickel from Nipron, or magnesium-deoxidized cathode nickels from titanium-deoxidized nickels. Spot tests have been very helpful in determining the presence or absence of nickel, chromium, and other elements. The spectrograph is also very useful for rapid identification of metals.

Corrosion Tests

Metals may be corroded in an accelerated manner by the spray or immersion method. A very common method is to use a spray of 5 per cent common salt solution on samples in a controlled-humidity chamber at 95 F. Immersion tests may be either total immersion or alternate immersion. For example, mercurous nitrate is commonly used for testing copper and its alloys (ASTM Spec. B154-51). Corrosion is generally measured by the loss of specimen weight in milligrams per square decimeter per day (mdd) or in average penetration in inches per year (ipy).

ELECTRICAL TESTING

Resistivity Checks

Resistivity checks on cathode strips or heater wire can identify the material or guide production control. Carbon losses in annealing cathode strips can thus be easily checked and regulated. The equipments consist essentially of a galvanometer, a battery, a wheatstone bridge, and a meter stick. By changing the variable resistance in the wheatstone bridge circuit to obtain balance, and by reading this resistance, the resistance of the specimen inserted as part of the wheatstone bridge circuitry is measured. The resistivity of the specimen is calculated from the resistance and the length and cross-sectional area of the specimen. Measuring resistivity at room temperature and at elevated temperatures gives the temperature-coefficient of resistivity or ratio of hot to cold resistivity, at different temperatures. This datum is necessary in the design of heaters.

Burnout Tests

Burnout data on wires or strips upon passage of electric currents are important in tube designs. The necessary equipment comprises binding posts for holding the specimens, a Variac transformer, a voltmeter, primary and secondary switches, and a transformer. By closing the primary switch and adjusting the Variac transformer, a desired voltage may be applied to the primary of the transformer and read on the voltmeter. The time required for the specimen to burn out from the instant the secondary switch is closed is a measure of the burnout resistance. This resistance is also affected by the surface emissivity and homogeneity of the specimen material.

A burnout test that indicates completeness of platinum coating is used in the testing of platinum-clad molybdenum grid wire.

MISCELLANEOUS TESTING

Weighing

Filament and grid data are generally based on accurate weighing of the component wire materials. These fine wires must be cut with clean ends to 200 ± 0.1 millimeters, and weighed on very sensitive balances accurate even to 0.1 milligram. From this weight reading, the average diameter of the wire can be calculated. Balances are periodically checked with check weights.

Powder Examination

The object of powder testing is to control the quality of a product, such as a heater wire or a ceramic part. There is often adequate correlation between some measurable size characteristics of the powders and desired quality. For most industrial purposes, three primary characteristics are measured: size frequency, specific surface, and particle shape. These characteristics are generally checked by the sedimentation, gas-absorption, and microscopic methods, respectively.

Surface Test for Identifying Metals

The surface test is used for rapidly identifying metals based on the following criteria:

1. Appearance, such as color, luster, crystallinity, smoothness, special surface marks, or fracture patterns

2. Blowtorch test--speed of melting, color change during heating, and action and appearance of slag

Spark Testing

By holding an unknown metal specimen against a grinding wheel with a constant pressure and observing the length of streaks, or the number, volume, and duration of sparks produced, the metal can sometimes be easily identified. Pure Armco iron is thus rapidly distinguished from steel, and steel from cathode nickel.

Color Comparison Test

Dumet lead wires and other materials have qualities closely related to their color. In the color comparison test for dumet, a range of quality standards is selected and is mounted against a black background. A sample is compared against the standards to determine which standard most closely approximates the color of the sample.

CATHODE MATERIALS FOR RECEIVING TUBES

I. S. Solet

The function of the cathode is to emit electrons. Because cathode materials emit more electrons when they are hot than when they are cold, the cathode is usually heated during operation. Heating may be done either by passing a current through the cathode (directly heated cathode), or by passing a current through a wire heater which the cathode encloses (indirectly heated cathode).

A cathode base material must satisfy three requisites:

1. It must be fairly plentiful and easily obtainable in a pure state.

2. The material should have certain physical requisites: high melting point, low vapor pressure, and high hot strength. It should be ductile enough to undergo routine forming operations. Its thermal emissivity should be low so that a minimum amount of heater power is required.

3. Among its chemical properties, the base material should be resistant to atmospheric oxidation so that it can be stored for long periods. Its oxide should be easily reducible in hydrogen so that a simple hydrogen firing can clean its surface. It should be fairly easily degassed. It should be stable at high temperatures and should not react undesirably with the oxide coating used to increase its electron emission.

The only material which will best satisfy these conditions is nickel. Iron, molybdenum, tantalum, and tungsten all have the required high melting point and low vapor pressure, but they react with the emissive oxide coating. The cost of platinum is prohibitive. Of course, for certain special applications, other cathode base materials may be used, but for most receiving tubes operating from an ac power source, the indirectly heated, oxide-coated, nickle-alloy base metal is the one most used. Although the following survey of cathodes will include all types, the greatest emphasis will be placed on the nickel-alloy-base-metal cathode.

DIRECTLY HEATED CATHODES

Directly heated cathodes are used: (1) with an ac source where it is not necessary to isolate the ac signal from the cathode (e.g., rectifiers), and (2) with a battery source where it is necessary to minimize current drain. It is also useful in equipment which must operate shortly after it is turned on since the directly heated cathode reaches operating temperature almost immediately.

Table V shows the mean efficiencies and operating temperatures of the different types of cathodes.

The higher efficiency of thoriated tungsten was discovered by accident when thoria was added to tungsten in an attempt to reduce grain growth which was causing rupture of the filament during operation. It was found that tungsten chemically reduces thoria to thorium metal which is a much more efficient electron emitter than tungsten. Both tungsten and thoriated tungsten are often used in high-voltage tubes where an oxide coating would be severely damaged.

Table V

Type of Cathode	Efficiency[1] milliamperes per watt	Temperature[2] deg C
Pure tungsten	1.7	2200-2500
Thoriated tungsten	6	1900
Oxide cathode, indirectly-heated	10	800-1150
Oxide cathodes, directly-heated	40	800-1150

The most commonly used filamentary cathode is the oxide-coated, nickel-alloy ribbon. Cobalt, tungsten, and aluminum have been added to increase the strength and electrical resistivity. Other elements such as carbon, magnesium, silicon, and titanium are reducing agents which are responsible for the production of barium. Manganese improves the metal-working properties. These metals will be discussed later.

Most nickel-alloy cathode ribbon is rolled from wire. The alloy is made by the vacuum melting and casting process which has proven very effective in controlling the composition to the necessary fine degree.

INDIRECTLY HEATED CATHODES

By far, the preponderance of receiving tubes uses the indirectly heated, oxide-coated cathode. This cathode is easily and inexpensively made. It is subjected to little or no mechanical stress and can be made thermally efficient by the use of a low-conductivity material for the tab which connects the cathode to an external terminal. Most important, the improved mechanical stability makes possible the close electrode spacing necessary for high gain; special nickels, in fact, may contain tungsten for increased strength. At the same time, ac hum is reduced by the shielding effect of the cathode sleeve.

Cathode sleeves may be made in the form of seamless (welded-seam) tubing or of lockseam cylinders. In both cases the nickel ingot is hot-rolled to about 1/8 inch strip and then cold-rolled. If seamless cathodes are to be made, the material is rolled to 0.020 inch, turned up and welded to form a cylinder which is then drawn to the desired size of seamless tubing. The tubing is then cut to cathode size, a tab is spot-welded to the end, and the sleeve is cleaned and coated. In some alloys (e.g. N18) seamless tubing is made by extruding a drilled ingot to an intermediate size tubing and drawing the tubing to finished size. Lockseam cathodes are made from strip which has been rolled and slit to finished size. The strip is fed into a cathode machine which blanks, embosses, and wraps the cathode to the proper dimensions. The tab may be blanked out with the cathode, in which case the assembly is called an integral-tab cathode; or the tab may be spot-welded to the cathode as in seamless cathodes. All cathode sleeves are made to extremely tight dimensional specifications so that consistently high performance may be attained. If the wall is too thick, operating temperature will be reached too slowly and there will be electrode spacing problems. If the strip for cathodes is not exactly the right size, the cathode-making machine will jam. Width tolerance of ± 0.002 inch and thickness tolerance of ± 0.0001 inch are specified. The strip should be clean to avoid problems of flaking of the coating from the cathode. Burr should be practically nonexistent so that the heater may be inserted with no trouble. The exterior surface finish of the cathode should be shiny to avoid radiation loss and to maintain a high cathode temperature for a given heater wattage.

GASES IN NICKEL

Gas is a very important consideration in materials used in a vacuum tube. A typical analysis of commercial nickel (99.14 per cent nickel) is given in Table VI.

Table VI[3]

Gas	Milliliters per 100 Grams Contained in Commercial Nickel
Hydrogen	0.43
Carbon Dioxide	1.81
Carbon Monoxide	0.22

It is thought that carbon monoxide and carbon dioxide do not exist as molecules in the nickel but rather are formed from dissolved carbon and oxygen which diffuse to the surface and combine. In any case, the removal of oxygen is a slow process, and it is better to control the amount of carbon and/or oxygen which is contained by the nickel. Hydrogen, on the other hand, diffuses quite rapidly through nickel, and thus, is an excellent gas to use for cleaning since it can be removed so easily.

"IMPURITIES" IN NICKEL AND THEIR RELATION TO EMISSION

Very early, it was discovered that oxide-coated cathodes made with very pure nickel sleeves were not efficient electron emitters. On the other hand, if one used commercially pure nickel, the emission was excellent. The improved emission resulted from the commercial practice of adding various deoxidizers and desulfidizers such as magnesium and silicon to the nickel while it was melted; these additions were responsible for activation and tube life. While initially this mechanism was not understood, today, the cathode core or sleeve is considered to be a reservoir containing reducing agents which diffuse to the coating where they reduce barium oxide to barium.

Table VII shows the variables needed to calculate diffusion coefficients in nickel according to the equation $D = D_0 \exp(-Q/RT)$, where D is the diffusion coefficient, D_0 is the proportionality constant, Q is the activation energy, R is the gas constant (1.987 calories/deg mole), and T is the absolute temperature.

To get some idea of the meaning of the diffusion coefficient, consider that the source of the diffusing species is infinitely large, and that the diffusing species is used

up entirely and immediately as soon as it gets to the coating; then, the amount q of diffusant arriving at the coating at steady state is expressed by

$$q = \frac{DC_o}{L}$$

where C_o = the constant concentration at the source end of the gradient in

$$qm/cm^3$$

L = the thickness of the cathode wall in cm

Table VII

Metal	D_o-cm^2/sec	Q-cal/mole	Reference Number
Aluminum	1.1	59,500	5
Magnesium	2.3×10^{-5}	31,300	6
Titanium	0.86	61,400	5
Manganese	7.50	67,100	5
Tungsten	11.10	76,800	5
Molybdenum	3.00	68,900	4
Silicon	10.6	64,800	6
Zirconium	1×10^{-5}	26,700	6
Nickel	1.27	66,800	5
Carbon*	0.1	33,000	7
Carbon*	0.456	39,330	8
Carbon*	2.48	40,100	9

*Results for carbon by three different experimental methods are given.

At cathode operating temperatures, carbon diffuses most rapidly, followed by zirconium, magnesium, aluminum, and silicon. Much work remains to be done on the effects of alloying interaction and grain size of the nickel base.

THE FUTURE OF THE THERMIONIC CATHODE

A discussion of cathodes would be incomplete without mentioning recent innovations aimed at increasing the life or the emissive power of the cathode:

1. The L-cathode is based on the reduction of a barium compound by porous tungsten. In one form, a BaSr carbonate reservoir is placed beneath a porous tungsten pellet.[10] In another, a tungsten plug is impregnated with barium aluminate. In a third, barium aluminate is mixed with tungsten powder and pellets are made out of the mixture. Instead of barium aluminate, barium silicate, titanate, thoriate, or tungstate may be used. Huber has made pellets out of a mixture of barium tungstate, aluminum powder and tungsten.[11]

2. J. M. Dodds has mixed barium carbonate with 1 per cent zirconium hydride (as a reducing agent) and nickel powder, and then made pellets out of the mixture, while A. H. Beck has used 0.1 per cent silicon instead of zirconium hydride.[11]

3. The lanthanum boride cathode is made by allowing lanthanum and boron to react in hydrogen at 1400 C, and painting or sintering the reaction product on a tantalum filament.[11]

4. The molded cathode (MacNair et al) consists of sintered mixtures of nickel powder and barium and strontium carbonates.[12]

Table VIII shows a survey of types of thermionic cathodes.

Table VIII[11]

Cathode Type	Operating Temperature—degrees	Power—watts/cm^2	Max. usable dc Emission—amp/cm^2	Life—hours	Work Function—e.v.
Tungsten	2600 K	80	0.5	10,000	4.5
Tantalum	2400 K	55	0.5	10,000	4.1
Thoriated Tungsten	2000 K	25	1-3	5,000	2.6
Lanthanum boride	1680 K	40	1	>250	2.6
Philips dispenser (Barium-Tungsten)	950 C (brightness)	4	1.3	8,000	2.1
Huber (Barium-Tungsten)	1050 C (brightness)	6	1.0	4,000	1.72
	1200 C (brightness)	8	4-5	400	
Dodds (Barium-Nickel)	950 C (brightness)	3.3	6	—	
Beck	1340 K	3.7	3	3,000	1.8
	1270 K	3.0	1.5	9,000	
Oxide Cathode	1100 K	2.8	0.5	3,000	1.2
	1200 K	4.2	1.0	1,000	

REFERENCES

1. Wagener, S., The Oxide-Coated Cathode, Vol. I, p. 7, Chapman & Hall, London, 1951
2. Basic Theory and Application of Electron Tubes, TM11-662, p. 13, United States Government Printing Office, Washington, 1952
3. Jackson, K. and R. O. Jenkins, "Nickel in Electronics," Metallurgia, 47, 277 (1953)
4. Peterson, R. W., D. E. Anderson, and W. G. Shepherd, "Influence of the Cathode Base on the Chemical Activation of Oxide Cathodes," J. Appl. Phys. 28, 22 (1957)
5. Swalin, R. A., and A. Martin, "Solute Diffusion in Nickel-Base Substitutional Solid Solutions," Trans. AIME, 206, 567 (1956)
6. Allison, H. W., and H. Samuelson, "Diffusion of Aluminum, Magnesium, Silicon, and Zirconium in Nickel," J. Appl. Phys., 30, 1419 (1959)
7. Gruzin, P. L., Yu. A. Polikarpov, and G. B. Fedorov, "An Investigation of the Diffusion of Carbon in Nickel and its Alloys by Means of Radioisotope C^{14}, "Fizika Metallov i Metallovedenie, 4, 94-102 (1957)
8. Code L1
9. Lander, J. J., H. E. Kern, and A. L. Beach, "Solubility and Diffusion Coefficient of Carbon in Nickel; Reaction Rates and Nickel-Carbon Alloys with Barium Oxide," J. Appl. Phys. 23, 1305 (1952)
10. Lemmens, H. J., M. J. Jansen, and R. Loosjes, "A New Thermionic Cathode for Heavy Loads," Philips Technical Review, 11, 341 (1950)
11. "New Forms of Thermionic Cathodes," Nature, 174, 1176 (1954)
12. MacNair, D., R. T. Lynch, and N. B. Hannay, "Molded Thermionic Cathodes," J. Appl. Phys., 24, 1335 (1953)

PLATE MATERIALS (METALS)

H. G. Scheible
T. A. Sternberg

The plate, or anode, of an electron tube is called upon to perform two basic functions: the collection of electrons from the cathode, and the dissipation of power caused by radiation and electron bombardment from the cathode. To perform effectively, plate materials must be good electrical conductors and should possess good heat-radiating properties. They must also be relatively low in gas content and should resist oxidation during fabrication and storage. Good ductility is also desirable since plates are formed into a variety of shapes and sizes. Mechanical strength is necessary if the plate is to retain its shape during cleaning, hydrogen firing, mount assembling, and high-frequency heating during exhaust.

Over the years, plates have been machined from solid graphite, pressed from carbon powers, and formed from metallic strip materials. Because metallic strip is the form most widely used today for receiving tube construction, this article will be confined to a discussion of that type.

Metals used in the manufacture of strip include: pure nickel, and its alternates—nickel-plated or clad steel, aluminum-clad steel, and a clad combination of aluminum on one side and nickel on the other side of steel. Two newer materials with special properties consist of a three-layer sandwich of copper, steel, and aluminum, and a five-layer composite of copper with aluminum-clad steel on each side. Molybdenum and tantalum are used as plate materials for special applications but will not be discussed here since they do not apply to receiving tubes.

Strip materials utilizing nickel, wholly or in part, are either clear (uncarbonized) or carbonized, depending on tube requirements.

Clear plates are easily degassed and are completely free of loose particles, which contribute to shorts and other difficulties in finished tubes. Their high reflectivity, however, usually causes increased cathode and grid temperatures, resulting in possible grid emission and leakage. Use of clear plates is generally restricted to tubes where heat dissipation per unit of plate area is low. The clear plate is preferred where the inside plate surface "sees" large voltage gradients.

Carbonized plates are more difficult to degas than clear plates and often show the presence of loose carbon particles. Their high thermal emissivity, however, is very advantageous and tends to suppress grid emission by lowering grid temperatures. Carbonized plates are used in tubes where dissipation per unit of plate area is high.

The aluminum-clad steel and aluminum-nickel-clad steel strips are low in gas, have good thermal emissivity after alitization (darkening of surface when heated above 658 C due to formation of the iron-aluminum intermetallic compound), and are free of loose particles. They are used in horizontal-and vertical-deflection tubes and other tubes where heat dissipation per unit of plate area is moderate and no loose particles can be tolerated. Since their thermal emissivities are lower than those of carbonized materials, other measures are sometimes necessary to compensate for this. For example, grid side rods are often made of better thermal conductors in an attempt to prevent overheating and grid emission. When substituting aluminum-clad strip for carbonized materials, the designer should be prepared to make adjustments to compensate for higher cathode and grid temperatures.

PROPERTIES OF COMMON STRIP MATERIALS

Clear Nickel, Nickel-Plated Steel, and Nickel-Clad Steel

In general, the ductility of clear materials is good and they can be readily fabricated into a variety of

shapes with minimum die wear. Weldability is usually good since oxidation does not occur for these materials under normal conditions. The slit edges of nickel-plated and nickel-clad steel must be protected by the use of specially treated wrapping papers to prevent rusting of the steel.

Nickel strip itself shows high dimensional stability at temperatures up to 900 C. The plated or clad materials, however, tend to warp and wrinkle at temperatures between 700 and 800 C, due to a phase transformation in the base steel. Distortion may sometimes be overcome by embossing or ribbing the plates.

Nickel, nickel-clad, and nickel-plated steel are easily degassed because of their relatively low gas contents and vapor pressures.

Nickel-plated and clad steels have good strength and maintain their shape well. Nickel strip, however, must be handled with care, since it does not have the structural strength of steel.

All three clear materials discussed are free of loose particles, but are poor thermal emitters.

Carbonized Nickel, Nickel-Plated Steel, and Nickel-Clad Steel

These materials are produced by either gas-carbonizing or by carbon-slurry carbonizing.

GAS-CARBONIZED MATERIALS

Gas-carbonized materials are produced by first oxidizing strip materials in air at approximately 925 C, and then firing in hydrocarbon gasses to form a fluffy, sooty layer of carbon. Excess carbon is removed by a brushing operation so that a thin, hard, shiny black surface remains. Brushing results in some loss of radiation properties, but is necessary to remove the loose mass of carbon formed in the process.

When nickel-plated steel (5 to 10 per cent total nickel) is gas carbonized, the strip may vary in hardness and springback due to carbon penetration of the steel core. Nickel-clad steel (20 per cent total nickel) also varies in springback, but to a lesser degree, since the heavier nickel layer probably tends to prevent carbon diffusion into the base metal. In general, gas-carbonized materials are fairly free of loose particles after brushing, but the thinness of the carbon layer limits hydrogen firing to 750 C.

SLURRY-CARBONIZED MATERIALS

In slurry carbonizing (an exclusive RCA process), clear strip materials are coated with either nickel oxide or carbonyl nickel and then hydrogen fired to form a highly porous sintered nickel coating. A mixture of carbon and graphite is then applied from a slurry containing nitrocellulose as a binder. After baking, the carbon coating is jet black, rough, and matte. This type of surface has excellent radiating properties and, since the carbon deposit is heavier than gas-carbonized coatings, may be safely hydrogen fired up to 850 C. Unfortunately, however, since the binder is removed during hydrogen firing, some loose carbon particles may be produced. Loose carbon may produce arcing and sputtering in tubes with closely spaced electrodes.

In general, strip materials of this type, prepared from carbonyl nickel, are the most satisfactory from the point of view of low gas content and freedom from impurities. Commercial nickel oxide contains trace quantities of sodium, zinc, lead, and chlorides which prove troublesome in tube manufacture. These impurities can cause undesirable bulb deposits and may contribute to cathode poisoning, high gas content, and leakage. In addition, complete reduction of the nickel oxide is never achieved during strip fabrication and traces of oxide remain, which give rise to initial and life gases.

Gas-carbonized or slurry-carbonized strips are more difficult to out-gas than clear strip. Welding on either type of carbonized strip does not present any particular difficulty. In general, welding currents are usually lower than for 100 per cent nickel strip.

ALUMINUM-CLAD STEEL

This material consists of a core of low carbon steel clad on both sides (by cold rolling) with aluminum (0.00045 to 0.00080 inch thick). The finished strip may have one or both sides brushed to remove excess oxide acquired during strip manufacture.

When parts are formed from this strip, the thin, hard surface layer of natural aluminum oxide causes a 25 per cent increase in wear on hardened steel dies. During hydrogen firing of fabricated plates, temperatures are normally held below 658 C to prevent alitization.

Thermal emissivity is good after alitization (which occurs during evacuation), but is lower than for carbonized materials. Gas content of this material is very low and it is free of loose particles. General physical characteristics resemble those of nickel-plated and nickel-clad steel. Welding is somewhat more difficult and requires higher currents, higher pressures, and more frequent maintenance of welding electrodes.

ALNIFER

Alnifer* consists of a low carbon steel base, clad on one side with aluminum (0.00045 to 0.00080 inch thick) and on the other side with a thin layer of nickel equal to 10 per cent of the total strip thickness. This material is also produced by a cold-rolling process. In general, its properties are similar to aluminum-clad steel, except that the exposed nickel surface remains shiny.

Alnifer is suggested for use in tubes where heater-cathode leakage problems are due to high heater temperature. The shiny, reflective nickel surface of the plate, facing the cathode, increases the cathode temperature and may permit redesign of a lower-temperature heater with reduced heater-cathode leakage. The clear nickel surface is free of loose particles. Recom-

* Trademark of Metals and Controls Division of Texas Instruments, Inc.

mended use is for high-voltage, close-spaced rectifiers. Welding problems are similar to those for aluminum-clad steel.

COPPER-BASE AND COPPER-CORED MATERIALS

Copper-base Aliron* is a three-layer material which consists of a sheet of steel, clad on one side with aluminum and on the other side with copper. It is recommended for use in rectifier tubes where the highest thermal concentrations are experienced.

Copper-cored Aliron is a five-layer material consisting of a core of copper, clad on both sides with aluminum-clad steel. Recommended use is for amplifier tubes where grid emission is a serious problem.

This family of materials combines the strength and stiffness of steel with the high electrical and thermal conductivities of copper and the self-darkening (alitization) character of aluminum. The steel is specially selected to have a recrystallization temperature below the alitization temperature. In addition, its low carbon content produces low springback in parts forming. The copper used is OFHC, which gives maximum electrical and thermal conductivity with low residual gas levels.

*Trademark of Metals and Controls Division of Texas Instruments, Inc.

The high conductivity of the copper produces more uniform heat distribution during exhaust, resulting in more effective degasification of the plate.

During tube operation, better heat distribution in the plate lowers over-all plate temperatures with a subsequent reduction in local overheating.

With copper-base Aliron, thermal deflection, or curvature, of the strip can occur as a result of differences in the thermal expansion of copper, steel, and aluminum. This tendency can be minimized by the use of known stiffening and reinforcing techniques in anode design.

This problem does not occur with the copper-cored material because of the symmetry of the sandwich layers. In general, welding is more difficult with these materials because of the high thermal conductivity and low electrical resistance of the copper.

Cost limits the use of copper-base and copper-cored Aliron to special applications.

PROPERTIES OF SOME TYPICAL PLATE MATERIALS

Table IX lists some typical plate materials together with the percentage of nickel plating or cladding, relative thermal emissivity, relative gas content, and relative cost.

Table IX

Properties of Some Typical Plate Materials

Type	Base Material	RCA Designation	Nickel Plating or Cladding—per cent per Side	Relative Thermal Emissivity*	Relative Gas Content†	Relative Cost Per Pound‡
Clear	Nickel	N7	-	0.35	10-20	185
	Nickel-Plated Steel	S79	2.5	0.35	10-20	44
	Nickel-Clad Steel	S104	10	0.35	10-20	160
Gas-Carbonized	Nickel	N61	-	0.80	150-200	314
	Nickel- Plated Steel	S139	2.5	0.80	150-200	181
	Nickel-Clad Steel	S102S	10	0.80	150-200	215
Carbon Slurry-Coated (Nickel Oxide)	Nickel	N48	-	0.98	175-225	239
	Nickel-Plated Steel	S8S1	5	0.98	300-350	89
	Nickel-Clad Steel	S105S	10	0.98	300-350	181
	Steel	S95D	-	0.98	300-350	35
Carbon Slurry-Coated (Carbonyl Nickel)	Nickel-Plated Steel	S311S4	5	0.98	200-250	100
	Steel	S310	-	0.98	200-250	-
Alitized	Aluminum Clad Steel	S72S1	-	0.85	10-20	80
	Aluminum-Nickel Clad Steel	S119	-	0.85	10-20	110
	Aluminum-Copper Clad Steel	C20S1	-	0.85	35-40	112
	Aluminum-Clad Steel on Copper	C11S1	-	0.85	35-40	155

* Emissivity is relative to black body.
† Gas content range in liter-microns per gram.
‡ Cost given in per cent of S311S4 strip price.

BIBLIOGRAPHY

1. Horsting, C. W., et. al., "Method For Determining Specific Cooling Rates of Electron-Tube Anode Materials in Vacuum," IRE Transactions on Electron Devices, January, 1959, 119-120
2. Technical Data Bulletin IND-4, Metals & Controls Corporation, General Plate Division
3. Espe, W., and E. B. Steinberg, "Aluminum-Clad Iron for Electron Tubes," Tele-Tech, February, 1951, 28-30
4. Haase, A. P., and E.B. Fehr, "New Vacuum Tube Materials," Tele-Tech, July, 1951, 33-35

GRIDS AND LEADS

A. J. Stoeckert

A grid is a tube element between the cathode and plate used to control the flow of electrons. Fig. 18 illustrates some typical grids.

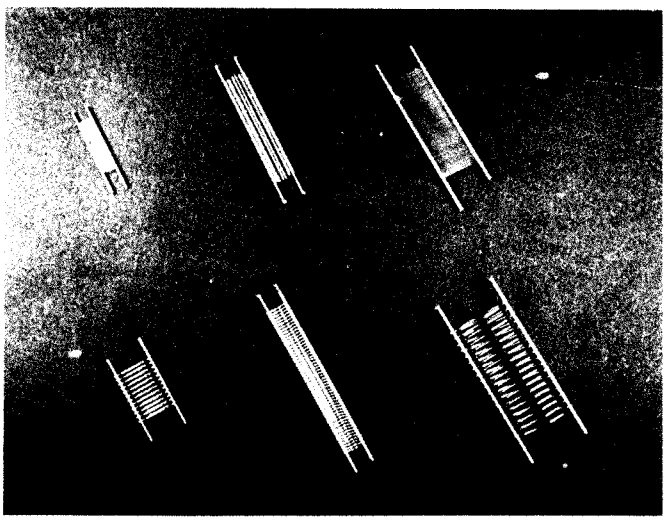


Figure 18. Typical Grids

CONSTRUCTION

Grids are constructed of fine wire on relatively heavy side rods. Some tubes have as many as five grids, one inside the other; all of the grids are similar in general form although different in dimensions. Grids are numbered in order from the cathode outwards, grid No. 1 being closest to the cathode.

Several types of grids are commonly used in receiving tubes; most are either lathe-wound or welded grids. In the past decade the welded grid, primarily because of its lower cost, has largely replaced the wound grid, except for control-grid use. Other types in commercial use are brazed cylindrical multisiderod grids and frame grids which will be discussed later.

Lathe-wound grids, as the name implies, are made on a lathe-type grid machine. This machine has a lead screw somewhat over a foot long. The lead screw, by its rotation, slowly pulls the two side rods through channels on each side of a rotating mandrel which determines the shape of the grid. As the side rods pass through the mandrel, the grid lateral wires are wound on it. A notching roller cuts a small V groove in the side rod and the fine lateral wire is wound in the groove; the side rod then rotates to a peening roller, which has a very narrow flat edge. The flat edge crushes the sides of the groove, thus clamping the grid lateral wire into the side rod. This process is illustrated in Fig. 19 for a right-hand-wound grid with the peening roller and the notching roller shown on the same side for simplicity.

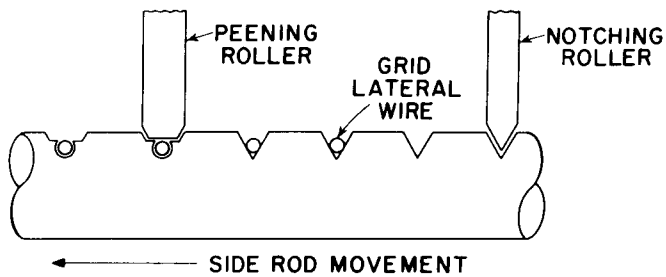


Figure 19. Lathe-Wound Grid Process (for a right-hand wound grid)

Some grids are left-hand wound, in which case the side rod movement and the position of the notching and peening rollers are reversed. Left-hand-wound grids are used when it is desirable to prevent line-up of grid laterals in a multi-grid tube.

At regular intervals along the grid strip, some turns are not peened but are allowed to rest loosely in the V grooves. The strip is cut into individual grids and the loose turns on each end of the grid are removed by hand. This is a very slow process, but is performed while the following grid strip is being wound. Newer grid lathes have devices to lift the notching roller in the leg portion of the grid. These "smooth-leg" grids along with pointed side rods (produced with special cutting tools) facilitate insertion of the grids into the micas during tube mounting.

Some grids may be completed after winding, while others may be sized in a separate operation. The most common wound-grid shapes are shown in Fig. 20.

Welded grids are made on an automatic machine (Fig. 21) into which the side rods are fed. The lateral wires are pulled across the side rods at the desired spacing. The laterals are all welded in place simultaneously. Most welded grids are sized and shaped, as they are made, on an expandable mandrel. A few are given further sizing or shaping treatments.

Welded grids cost less than lathe wound grids, but the finest lateral wire that can be handled is 0.002-inch in diameter. Wires finer than 0.002-inch break too easily when pulled through the machine. Spacing on

welded grids is not as accurate as on lathe wound grids. Therefore, welded grids are generally limited to 100 to 110 turns per inch (T.P.I.). Provided spacing is not critical, as many as 125 or 140 T.P.I. may be used. Another limitation for welded grids is that high thermal conductivity materials, such as copper and copper alloys, cannot be readily welded. For these reasons, No. 1 grids are generally lathe wound; the others may be welded. The most common welded grid shapes are shown in Fig. 22.

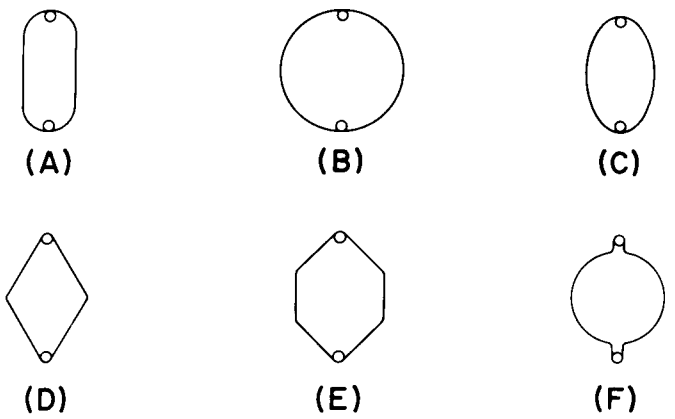


Figure 20. Common Wound-Grid Shapes: (A) flat; (B) round; (C) elliptical; (D) diamond; (E) hexagon; (F) formed round

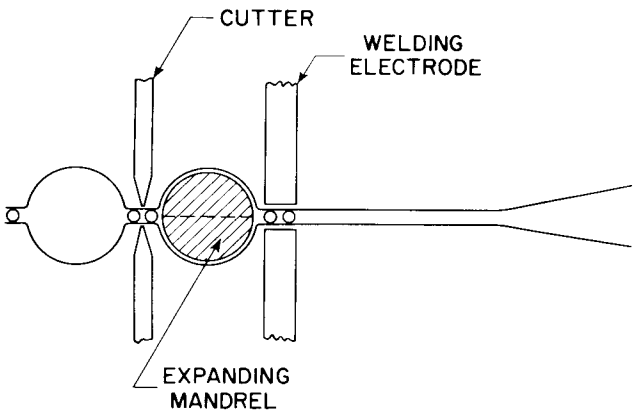


Figure 21. The Welded-Grid Process

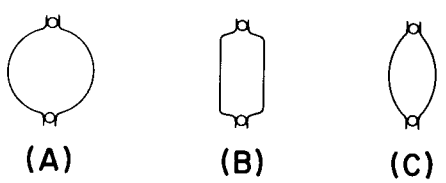
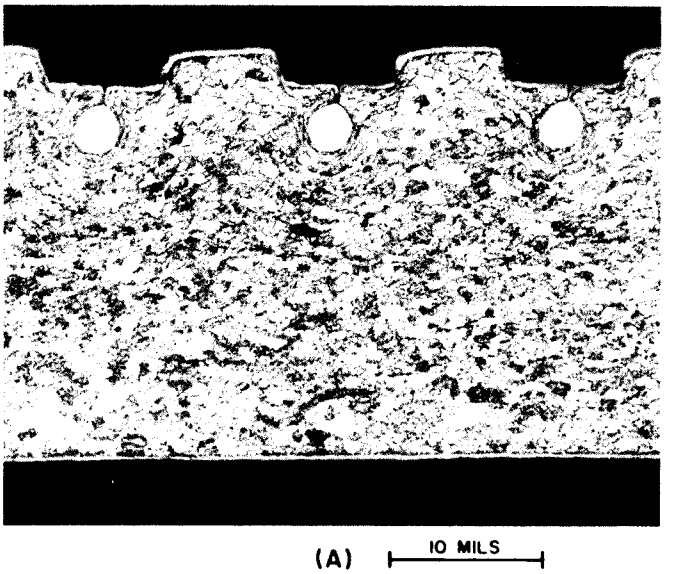


Figure 22. Common Welded-Grid Shapes: (A) round; (B) flat; (C) elliptical

A comparison of the structure of lathe wound grids and welded grids is illustrated in Fig. 23. Fig. 23A is a photomicrograph of a longitudinal section through a Nipron side rod of a lathe wound grid showing the lateral wires peened in place. Fig. 23B is a photomicrograph of a longitudinal section through a Nipron side rod of a welded type grid. The fine lateral wires are shown flattened and welded to the side rod.



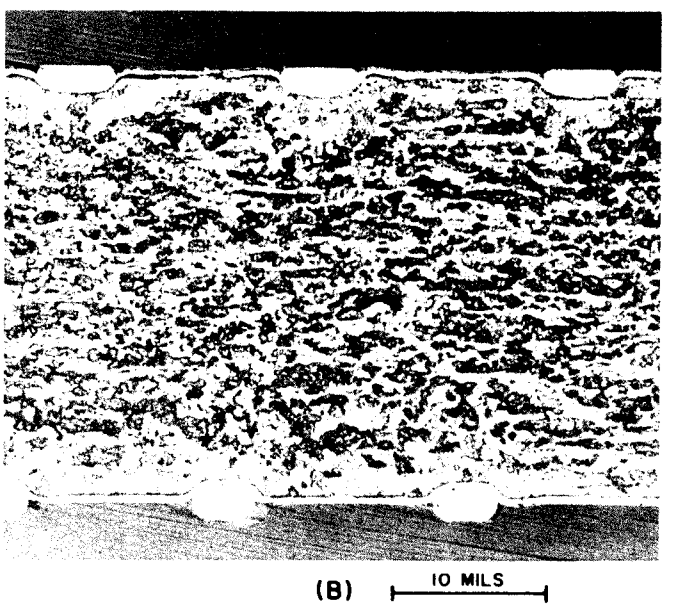


Figure 23. Longitudinal Sections of Grid Side Rods at Lateral Wire Connections: (A) lathe-wound grid; (B) welded grid

Materials — Lateral Wires

Materials for grid lateral wire should have the following properties:

1. High thermal conductivity to minimize temperature rise and keep the grid cool enough to prevent thermionic emission
2. High work function to discourage grid emission
3. Low secondary emission
4. Proper thermal expansion (usually low) to keep

distortion to a minimum during the processing and life of the tube

5. High hot strength to prevent sagging at the high tube-processing and operating temperatures

6. Uniform tensile strength and elongation; uniform elongation is desirable during the manufacture and sizing operations

7. Good weldability (for welded grid making)

8. Wire should draw easily to fine sizes and cannot be out-of-round or uneven spacing will result

Molybdenum most closely fits the requirements for grid lateral wire and for many years was the most common grid material. Now, however, due to the relatively high cost of molybdenum, alloys of nickel have come to the fore. As a matter of fact, 5 per cent manganese nickel (Magno[1] or D Nickel[2]) is used for most lateral wire for receiving type tubes. For grid wire below 1.0 mil in size, tungsten is generally used because of its high tensile strength. Grids requiring particularly high hot strength are made with Hastelloy[3], molybdenum, dowmo, or tungsten. Because of their close proximity to the cathode, No. 1 grids operate at a high temperature. However, No. 2 grids, with higher positive voltage and more electron bombardment than No. 1 grids, are very often hotter than No. 1 grids. Grid emission is not a problem with screen grids because of the low circuit impedance of screen-grid circuits. Control grid circuits usually have high impedance making them very sensitive to grid emission. For these reasons lateral wire on No. 1 and No. 2 grids are often molybdenum or tungsten; they may be plated with a noble metal to minimize grid emission.

Materials recently developed for grid lateral wire include palladium-plated wire and gas-carbonized wire. Palladium has a lower vapor pressure than silver or gold and therefore, can be run hotter, allowing higher cathode temperatures and better outgassing of tube elements. Carbonized wire has been developed as a substitute for carbon-sprayed grids. The rough black surface is an excellent heat radiator so that grid wire runs cooler in most tubes and has lower primary and secondary emission.

Materials — Side Rods

Materials for grid side rods should have the following properties:

1. High thermal conductivity to minimize temperature rise and thus to prevent the grid from becoming an emitter

2. High stiffness to hold the grid laterals rigidly in place

3. Good weldability to lateral wires and to the stem

Pure copper has the best conductivity, but, unfortunately, is too soft and is difficult to weld. Small additions of chromium, cadmium, or silver to copper improve its stiffness and weldability without having a serious detrimental effect on its thermal conductivity. Chrome-copper, which has 1 per cent chromium, is the copper alloy most often used. No. 1 grids usually have chrome-copper side rods because they are the most critical for grid emission and hence temperature. Where the temperature is not as critical, grids have nickel-plated iron or copper-clad iron for their side rods.

Those metals and alloys most frequently used for grids are listed in tables X, XI and XII. It should be kept in mind that any of the lateral wires may be plated with silver, gold, palladium, or platinum. These data were taken from a summary of grid wire properties by J. J. Carrona, E. R. Schrader, and D. Yanchusk.

Multisiderod grids, as used in nuvistors and pencil tubes, have from 12 to 70 side rods. Such grids are made on a grid lathe especially constructed for the purpose. The side rods (fine wire from under 1 mil to 6 mils) are fed through a nose cone which spaces them evenly around a mandrel upon which the grid is supported while being wound and brazed. The lateral wire is wound over the side rods at the desired T.P.I. Grid lengths wound on mandrels about 10 inches long are brazed by passing through a high-frequency coil in a hydrogen atmosphere. The temperature is raised to melt the copper plate on the lateral wire. The copper flows to and alloys with the nickel plating on the side rods and a mesh structure is formed. The grid is then slid off the mandrel and cut into length on a cutting mandrel.

LEADS

Leads are wire elements which connect the tube elements with the external circuit. They differ depending on the tube type: miniature, metal, or glass. Most leads are composed of three portions butt-welded together. These portions are: (1) the inner lead, which is inside the tube and makes contact with the tube elements; (2) the seal portion, which contacts the glass and forms therewith a vacuum-tight seal; and (3) the outer lead, which contacts the base pin or socket.

The sizes and most common material used for the various types are given in Table XIII.

The three-piece lead may occasionally be reduced to a two-piece lead or even to a one-piece lead. In the one-piece lead, the entire length is of the same material; in a two-piece lead, the outer lead and seal lead are the same. An essential requirement of a lead is that it make a vacuum-tight seal with the glass. In a glass stem, the expansion of the lead material must match as closely as possible the expansion of the glass. In a ceramic stem, the expansion of the lead plus that of brazing material must approximate the expansion of the ceramic. Since compression seals are stronger that seals in tension, it is advisable to have the lead and braze combination of slightly higher expansion than the ceramic expansion. (See articles on glass-to-metal and ceramic-to-metal seals.)

[1,2,3] Superscript numbers refer to "List of Manufacturers of Alloys" at end of section.

Table X

Properties of Materials Used in Grid Manufacture

Material	Relative Cost	Composition	Tensile Strength pounds/in^2 x 10^{-3} 20 C	500 C	800 C	Elongation max.	Resistance micro-ohm-cm. at 20 C	Thermal Conductivity Watts/cm/deg C
Lateral Wires								
Magno[1]	1.0	5% Mn-Ni	86	50	20	40	20	0.481
Permanickel[2]	1.4		100	60	22	40	16.6	0.587
Stainless Steel	1.7	18% Ni, 8% Cr	150	65	20	30	72	0.133
Hastelloy A[3]	1.8	Ni, Mo, Fe, Mn	120	90	47	48	126.7	0.17
Hastelloy B[3]	2.4	Ni, Mo, Fe, Mn	140	115	60	45	135.0	0.115
Gridnic C[1]	2.7	15% Cr-Ni	90			30	100	1.38
Molybdenum	8.9		150	87	63	30	5.3	1.46
Tungsten	13.0		500	175	130	2	5.5	2.0
Dowmo4	13.1	50% Mo, 50% W	185		110	20	9.0	
H Wire[4]	17.1	80% Mo, 20% W	175			25	7.6	
W Wire[4]	20.9	75% Mo, 25% W	180			25	8.1	
N-100[4]	26.0	2% W, 1% Al-Ni			25	6	17	
Side Rods								
Iron	0.5		45	27		40	10.5	0.745
Nipron	1.0	Ni-plated Fe	47			40	10.5	0.725
Copperweld[5]	1.4	30-40% Cu-clad Fe	41			25	3.5	1.98
OFHC Copper[6]	1.6		35	13	12	50	1.76	3.94
Chrome Copper	2.2	1% Cr-Cu	36	23		35	2.30	3.94 approx.
Nickel	2.5		77	50	22	45	9.4	0.603
Kulgrid[7]	4.2	29% Ni-clad Cu	48			45	4.3	2.92

[1-7]Superscript numbers refer to "List of Manufacturers of Alloys" at end of section.

The seal-lead material used in most receiving tubes with soft glass stems is called dumet. Dumet is a composite wire with central core of 42 per cent nickel-iron alloy and an outer sleeve of pure copper. The copper sleeve comprises about 20 per cent of the total cross-sectional area. The expansion of dumet is not the same in the radial direction as it is in longitudinal direction. The radial expansion is 10.25×10^{-6} inches per inch per deg C and the longitudinal expansion is 7.6×10^{-6} inches per inch per deg C. Because of this difference, it is desirable to keep the dumet section of the lead, which is embedded in the glass, short.

The purpose of the dumet section is to form a vacuum-tight seal with the glass stem. Although nickel-iron alloys of about 48 per cent nickel match the expansion of glass, they do not make good bonds to glass and their electrical resistance is high. However, copper, or rather copper oxide, makes a good bond with glass, and copper has excellent electrical conductivity. Thus the idea of dumet came into being when a sleeve of copper was bonded to a nickel-iron-alloy core. With the increased expansion of the added copper, the nickel-iron alloy was changed to a lower-expansion 42 per cent nickel-iron so that the composite would approach the desired expansion of soft glass.

Table XI

Lateral Wire Material	RCA Spec.No.	Relative Cost Per Unit Length (basis Magno = 1.0)	Relative Use—per cent of total
Magno[1]	N13W	1.0	50
palladium plated	N24W	12.0	
silver plated	N27W	1.4	
gold plated	N28W	12.0	
Molybdenum	M13W	8.9	34
gold plated	M34W	19.5	
Permanickel[2]	N123W	1.4	7
Tungsten	T10W	13.0	4
gold plated	T39W	29.5	
silver plated	T40W	29.5	
Dowmo[4]	D18W	13.1	5
palladium plated	D2W	23.7	
gold plated	D30W	27.6	
Hastelloy "B"[3]	H26W	2.4	
gold plated	H28W	13.0	

[1-4]Superscript numbers refer to "List of Manufacturers of Alloys" at end of section.

Table XII

Side Rod Material	RCA Spec. No.	Relative Cost Per Unit Wt. (Basis Nipron = 1.0)	Relative Use—per cent of total
Nipron[4]	I23W	1.0	80
	I29W	1.1	
	S305	1.1	
Chrome Copper	C92W	2.2	16
Copperweld[5]	S148W	1.4	4
Nickel	N5W	2.5	

[4,5]Superscript numbers refer to "List of Manufacturers of Alloys" at end of section.

To further improve the bonding of the dumet with the glass and to prevent excessive oxidation of the copper during stem making, the surface of the copper is coated with borax. The color of dumet may range from a light straw color to a dark red, depending on the thickness of oxide on the copper. A usable range of oxide thickness on dumet as received is from 0.03 to 0.15 mil; after sealing it should range from 0.03 mil to 0.3 mil. If it exceeds 0.3 mil, there is danger of slow leaks through the porous or cracked oxide layer. Oxides as heavy as 0.5 mil are certain to cause leaks. On the other hand, an absence of any oxide at all is sure to produce a leak as well as a weak seal.

Table XIII

Lead Materials

Tube Type	Inner Lead	Seal Lead	Outer Lead
Miniature	0.020 Nipron[4]	0.016 Dumet[8]	0.040 nickel
	nickel on premium		0.040 Ni-plated Copperweld[5]
Metal	0.020 Nipron	0.016 Dumet	0.020 copper
Glass	0.020 nickel	0.016 Dumet	0.020 nickel
	0.020 Nipron	0.020 Dumet	0.020 Copper-weld
	0.030 or 0.040 Kulgrid[7]	0.025 Dumet	0.040 nickel
			0.040 Ni-plated Copperweld
Subminiature glass	0.017 Dumet (one-piece lead)		
Nuvistor	0.0155 molybdenum (one-piece lead)		

[4,5,7,8]Superscript numbers refer to "List of Manufacturers of Alloys" at end of section.

In the two-piece lead, dumet forms the outer lead and the seal portion, while most one-piece leads are made entirely of dumet. However, since the cost of dumet is relatively high and since dumet does not fulfill all the requirements for the inner and outer leads, a three-piece lead is generally used. Most leads are composed of a short piece of dumet in the center, butt-welded to an inner lead and an outer lead of different materials.

Since inner leads make the connections to the tube elements, they must have good welding properties and must be soft enough to bend into the desired configuration for connection to the tube elements. Annealed nickel is used, with nickel-plated iron as a logical substitute.

As mentioned before, outer leads differ with the tube type and socket requirements. Miniature tubes have a stiff outer lead, or pin, which is usually of 0.040 nickel wire. The pin must be stiff enough to resist excess distortion, yet soft enough to yield to the socket requirement without breaking the glass seal. Stiffness should be between 30° and 60° bend with a 550-gram load on the RCA 719H stiffness tester.

Metal tubes have 0.020 copper outer leads which are fed into the base pins and then clipped to the length of the pins. They are soldered to the base pins by fluxing and dipping into a solder pot. The copper lead wires must be flexible enough so that they can be bent to shape and fed through the base pins.

There are two types of outer leads on glass tubes: 0.020 inch flexible leads and 0.040 inch stiff leads. Both are soldered into base pins just as for the metal tubes.

The heavy stiff leads are used where much heat is dissipated in the tube and where lead conduction is relied upon to carry part of this heat away. The butt welds in two- and three-piece leads must be strong enough to withstand bending of the lead 90° and back through 180°. The weld knot must also be free from rough and ragged edges, points, or other irregularities that might be starting points for cracks in the glass stems, and might prevent feeding in the stem machine funnels.

LIST OF MANUFACTURERS OF ALLOYS

1. Driver Harris
2. International Nickel
3. Haynes Stellite
4. RCA
5. Copperweld Steel Co.
6. United States Metals Refining Co.
7. Brookside Metal Co.
8. Westinghouse

METAL-TO-METAL JOINING

E. S. Thall

Although the majority of metal-to-metal bonds or joints in electron tubes are made by electric resistance welding, close inspection of a tube will reveal at least one other technique currently in use. According to the method of manufacture, joining of tube parts may be classified in four groups:

1. Electric welding
2. Brazing and soldering
3. Cold Welding
4. Mechanical joining: riveting, folding, notching and peening

In general, for a good metal-to-metal joint, freedom from oxide, dirt, or grease is of prime importance; it is an absolute necessity if the parts are to be joined by brazing or soldering. Cleaning processes commonly used are vapor and solvent degreasing, ultrasonic cleaning, hydrogen firing, and mechanical abrading (barrel cleaning, polishing, brushing), or combinations of these methods.

ELECTRIC WELDING

Three types of welding are commonly used: (1) spot welding, (2) projection welding, and (3) butt welding. All are forms of resistance welding, in which the heat required to fuse the base metals is developed by the flow of electric current through the resistance offered by the materials being joined. Of the three types of welding mentioned above, the spot welding technique is the simplest, and the one most widely used in receiving tube manufacture. The small bench welders used for this purpose are 1.5 kilovolt-ampere machines and deliver up to three volts on the secondary, or welding, circuits. Since the resistance across the weld is small (usually a fraction of an ohm), the currents are correspondingly large (up to 500 amperes). A stepdown transformer provides the low voltages and high currents necessary. The following description of the technique of spot welding, while applicable to spot welding in general, has been written with the bench welders now in factory use in mind and, as such, is specific to this unit. (See Figs. 24 and 25).

To make a spot weld on one of these machines, the operator holds the two pieces being joined in alignment and in contact between the electrodes. Pressure is then applied to the electrodes through a linkage from a pedal and, as the pedal is depressed, the load on the work builds up to a preset maximum (usually about two pounds). The application of the load is followed by the throwing of a switch which permits the current to flow. The heat developed at the point of contact of the parent metals causes the parts to fuse locally which, along with the pressure being continuously applied, results in the formation of a "bridge," or "nugget," between the parts. The nugget has, for practical purposes, the same composition as the parent metals. Three variables are involved in the process: the force or mechanical pressure applied to the electrodes, the magnitude of the current, and the duration of the current flow. A transformer is normally used to supply the high-current, low-voltage power required. The duration and magnitude of the current are controlled in the primary circuit. The duration of the current pulse is less than one second, and is controlled electronically. The control timer may be either synchronous or nonsynchronous. The synchronous timer is used in critical applications and always starts the timing when the instantaneous voltage is zero. Precise control — down to a portion of a cycle (at 60 cycles per second) — may be achieved in this manner.

The magnitude of the current is controlled by adjusting the primary voltage or by using a phase-shift control thyratron, and varying the conduction angle.

When the current is regulated by a timer, synchronous or otherwise, the welder is called a timed ac welder. Another type of welder currently finding increasing favor is known as a capacitor-discharge welder. This welder employs a bank of capacitors in the primary circuit; the magnitude of the current is controlled by the number of capacitors and the voltage to which they are charged. The discharge time is determined by the characteristics of the discharge circuit. The discharge takes place, as in the timed ac welder, when the switch is closed as the weld is about to be made. The current is dc, and because the welding time is of the order of microseconds, the currents are very large. The very large current passing for a short time is especially useful for welding: (1) metals subject to oxidation, (2) small parts made of high melting point metals, such as tungsten and molybdenum, and (3) very low-resistance metals such as copper. With low-resistance metals, the large currents develop enough heat to fuse the metal. Since the welding time is very short, the high thermal conductivity of the base metal does not seriously interfere with the process. On the other hand, oxidized surfaces can cause considerable difficulty with capacitor-discharge welders because most of the heat generated locally at the interfaces causes overheated welds with

much splashing. For materials that are subject to oxidation on the surface, the timed ac welder often proves more suitable because the oxide can be broken down slowly, and there is time for the heat to be conducted away to the parent metals. In either the capacitor-discharge or timed ac welder, the mechanical load, or pressure, on the electrodes is varied by changing the spring tension or compression as it is applied through the pedal. Although the usual pedal loading for a bench welder is about two pounds and may go as high as ten pounds, the pressure on the weld (pounds per square inch) is many times this as the area in contact is but a small fraction of a square inch. The heat developed in a weld is dependent on the resistance of the materials being welded as well as on the magnitude of the current. The heat developed depends on the I^2R relationship; if a large resistance should occur between the electrodes and the base metals, considerable heat will be developed at these interfaces, with resultant sticking of the electrodes. Good conductivity must be maintained at the electrode-to-base-metal interface which should be quite clean and free of extraneous dirt and oxide. The electrodes themselves should have as low a resistance as possible to minimize sticking. Special electrode alloys have been developed which combine the high electrical conductivity of copper and silver with hardness of materials such as tungsten, to produce an electrode with long life and good welding characteristics. Excessive heat developed at the weld itself can result in the fusion of an excessive amount of metal. The attendant large sputter or splash may result in an oxidized or burnt condition in the nugget and a characteristically brittle weld. In initial set-up for a job, once a good weld has been obtained, the pressure on the electrodes should be reduced as far as possible without incurring splashing or sticking in order to produce minimum electrode distortion and wear and maximum electrical efficiency.

It is important to keep the electrical input as low as possible so that: (1) the area of the heat-affected zone can be kept to a minimum, and (2) the weld can be kept ductile and can retain properties similar to those of the parent metals. The time should be reduced first so that the possibility of oxidation with resulting embrittlement will be minimized. The current may have to be increased to compensate for the reduced time. Reducing the pressure, although it decreases the current through the weld (since the resistance is increased) may result in more splash and more sticking with increased local heating at the metal interfaces. Consistent with good fusion, little splash, and no sticking, the pressure on the electrode should be kept to a minimum. Fig. 26 shows cross-sections made through five projection spot welds. While projection welds are, in fact, dealt with in Fig. 26, the method of making a good weld by the spot welding technique is so closely parallel to this that the considerations involved apply equally to both. The welds illustrated are the main-seal welds in a metal tube. A of Fig. 26 shows a cold weld where the pressure and input are too low; B shows a cold weld where the input is sufficiently high, but where the pressure has been increased to the point where the resistance and, hence, heat developed at the interface of the parent metals is too low, resulting, likewise, in a cold weld; in C the input is sufficiently high, but the pressure has been reduced to the point where considerable local heating has been obtained with resultant splashing; D shows a satisfactory pressure, but the input is so high that the weld has been embrittled by overheating resulting in characteristically large nugget and coarse grain structure; E shows a proper weld with the correct pressure and input: the parent metals are completely fused, but the fine grain structure of the nugget has been preserved.

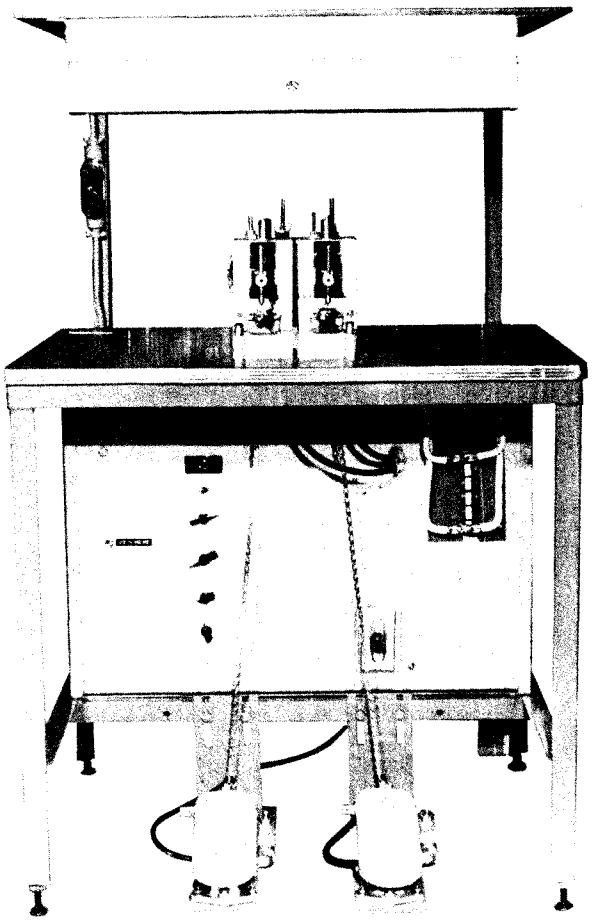


Figure 24. Bench Welder

As was mentioned earlier, beside the spot-welding technique dealt with above, two other methods, namely butt and projection welding, are used. Seam welding, where a series of spot welds is made to produce a single long continuous weld, may be considered as being merely a variation of the spot-welding technique. Projection welding is similar to spot welding. The main difference is that the base metal is raised, or embossed, to localize and concentrate the current and to increase the resistance so that a weld can be made within the current capabilities of the welder. It also permits uniform welding on broad surfaces, which otherwise would tend to result in uneven heating causing a nonuniform weld. Projection welding is used for making the main seal in metal tubes and for grid discs in pencil tubes, to mention only two of many applications.

Butt welding, on the other hand, is somewhat different from either spot or projection welding, although it too is of the resistance type. The pieces to be welded are fixed firmly in end-to-end contact in an electrode clamp

or vise. The materials are then heated to near fusion by the passage of an electric current. At this instant, the parts are forced together under high pressure and the weld is made. Butt welding is not used for tube assembly as are spot and projection welding; it does find use, however, during the drawing operation of nickel and nickel-alloy wire and for dumet leads where the dumet section is butt welded to the inner and outer leads.

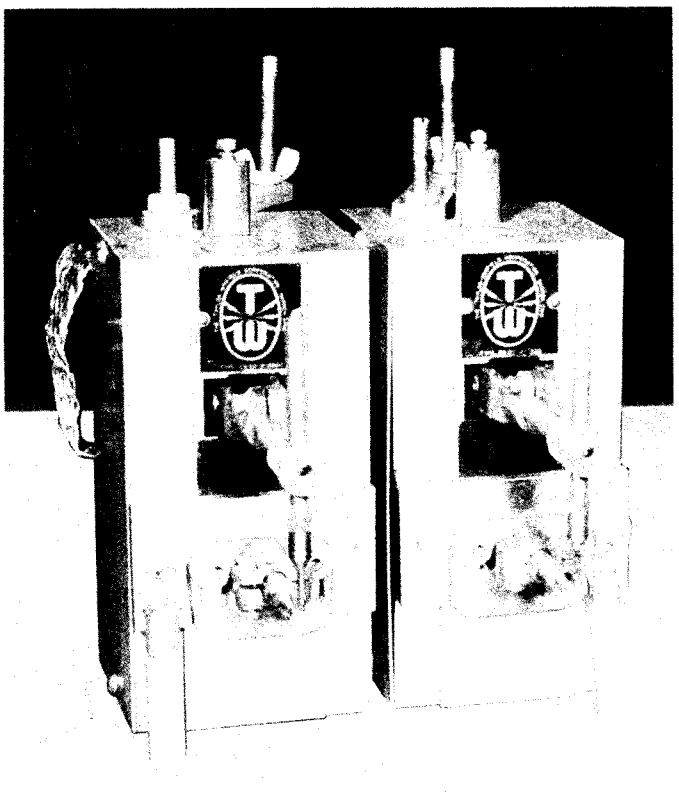


Figure 25. Welding Heads of Bench Welder

While most materials, or combinations of materials, can be joined by the timed ac spot-welding technique mentioned above, a few are difficult, if not impossible, to weld. These materials are those with high melting points, such as tungsten and molybdenum and their alloys, and the high conductivity materials, such as copper and silver. Metals and alloys of the tungsten and molybdenum group have very high melting points, good conductivity, and are easily oxidized; metals and alloys of the copper and silver group have not sufficient resistance to develop the heat needed to make a good weld. It is understandable why copper and tungsten metals and alloys of them form the basis of most electrode materials. P. R. Mallory and Company make a series of alloys known as Elkalloy and Elkonite specially for welding electrodes. The compositions of electrodes most used in the receiving tube operation are given in Table XIV.

In the automatic grid-welding (Autoweld) machine the electrodes are tipped with tungsten to guard against excessive wear caused by the high pressure (about 100 pounds load) needed for welding the side rods to the laterals. The automatic welder makes a large number of spot welds to simultaneously join all lateral wires to the side rods in one motion of the weld head.

Table XIV

Welding-Electrode Alloys

Alloy Name	Composition per cent
Elkalloy A	99 copper - balance cadmium
Elkonite 10W3	50 copper - 50 tungsten
Elkonite 20W3	40 copper - 60 tungsten
Special Elkonite 100W	20 silver - 80 tungsten
Mallory 3	97 copper - 3 cobalt + beryllium
Mallory 100	99-1/2 copper - 1/2 cobalt + beryllium

BRAZING AND SOLDERING

Brazing and soldering are used (with the exception of the grid braze) externally for radiation fins, for joining outer leads to socket pins and top caps, and for exhaust tubulations where the braze forms a hermetic seal between the inner wall of the electrode and the external tubulation. Brazing and soldering differ from resistance welding in that a second, or filler, metal is introduced between the parts being joined. The filler metal (braze or solder) fuses on heating and runs over the surface of the parts wetting them so that they bond on cooling. Although no clear-cut distinction exists between brazing and soldering, the latter refers usually to the operation when it is performed at low temperature, technically, below 800 F. The filler metal must always melt and flow at a temperature below the melting point of the materials being joined. Cleanliness of the parts is most essential if a good joint is to be obtained. Either chemical or combinations of chemical and mechanical cleaning may be used. In the brazing operation, a protective atmosphere of hydrogen may be used; for the lower temperature used in soldering, a chloride flux only is necessary, and the operation is usually carried out in air. In the pencil tube and nuvistor, the brazes must be exceptionally free of contamination and for this reason no zinc or cadmium can be tolerated in the braze metal. Common materials used for making brazes and their flow points are shown in Table XV.

Brazing is used extensively in pencil tubes and nuvistors for such purposes as sealing-in the ceramic stems, securing grid lateral wires to side rods, connecting the copper exhaust tubulations to the terminals (pencil tube), and connecting the internal elements to the leads (nuvistor). In the brazing of the side rods and laterals in the pencil tube, the gold-plated tungsten lateral wire is brazed to the silver-plated copper side rod wire; the gold, the silver, and possibly, some of

the copper form the braze material. No separate filler metal is used for the grid braze. The grid and exhaust tubulation brazes in the pencil tube are made by induction heating; the operation is carried out in a hydrogen atmosphere. It should be pointed out that intergranular penetration of the braze into the base metal, with subsequent embrittlement and porosity, is an ever-present danger if high silver brazes are used on iron-nickel alloys, for example, Kovar. The difficulty may be overcome by first plating with nickel to a thickness exceeding 0.00025 inch. The hazard can be reduced further by using minimum brazing time.

Table XV

Typical Brazing Metals and Alloys Used in Electron Tube Manufacture

Alloy	Composition per cent	Melting Point deg C	Flow Point deg C	Flux	Remarks
Gold-Indium	80 Au, 20 In	550	630	Hydrogen	Powder
Easyflo*	50 Ag, 16 Cu 16 Zn, 18 Cd	627	635	Handy Flux	
Silfos*	80 Cu, 15 Ag 5 P	641	704	None	
RT*	60 Ag, 25 Cu 15 Zn	686	718	Handy Flux	
BT	72 Ag, 28 Cu	779	779	Hydrogen or Handy Flux	Ag-Cu eutectic
BTL	60 Cu, 40 Ag	778	910	Hydrogen or Handy Flux	
Gold-Copper	80 Au, 20 Cu	890	890	Borax** or Hydrogen	Lowest melting point gold-copper braze
SN7*	85 Cu, 7 Ag 8 Sn	946	985	Borax or Hydrogen	
Gold-Copper	37 Au, 63 Cu	960	1000	Borax or Hydrogen	
Silver	100 Ag	960	960	Borax or Hyrdrogen	
Gold	100 Au	1063	1063	Borax or Hydrogen	
Copper	100 Cu	1082	1082	Hydrogen	
Nickel	99 Ni + Co	1450	1450	Hydrogen	Wets Mo and W
Platinum-Cobalt	60 Pt, 40 Co	1450	1450	Hydrogen	Wets Mo and W
Platinum	100 Pt	1771	1771	Hydrogen	Wets Mo and W

*Handy & Harman Co.

**Borax may be applied as dry powder or as a paste of 10 per cent boric acid in borax

Low-temperature soldering is carried out in air, and for this, the use of a flux to obtain proper wetting, or tinning, of the surface by the solder is necessary. Socket pins and top caps are joined to the outer leads by soldering. At present, pure tin (RCA S-5) or a 35 per cent tin - 65 per cent lead (RCA S-28) solder is used for the socket base pins, and either the RCA S-28 or a 97 1/2 per cent lead - 1 per cent tin - 1 1/2 per cent silver (RCA S-136) is used for soldering the top caps. For the top caps, the solder is used in the form of wire.

For other applications, the parts are dipped first into an aqueous solution of ammonium and zinc chlorides for fluxing and then into a pot of solder.

COLD WELDING

In addition to the methods of fusion by the application of heat, as in electric welding, brazing, and soldering, another method, that of cold welding must be mentioned. In cold welding, although no heat is applied to the joint, a bond is made at room temperature which is hermetically tight and which is seen, microscopically at least, to be completely fused. The technique consists of simply forcing the parts together under very high pressure; its use, presently, is confined exclusively to the soft metals such as copper and aluminum. For the pencil

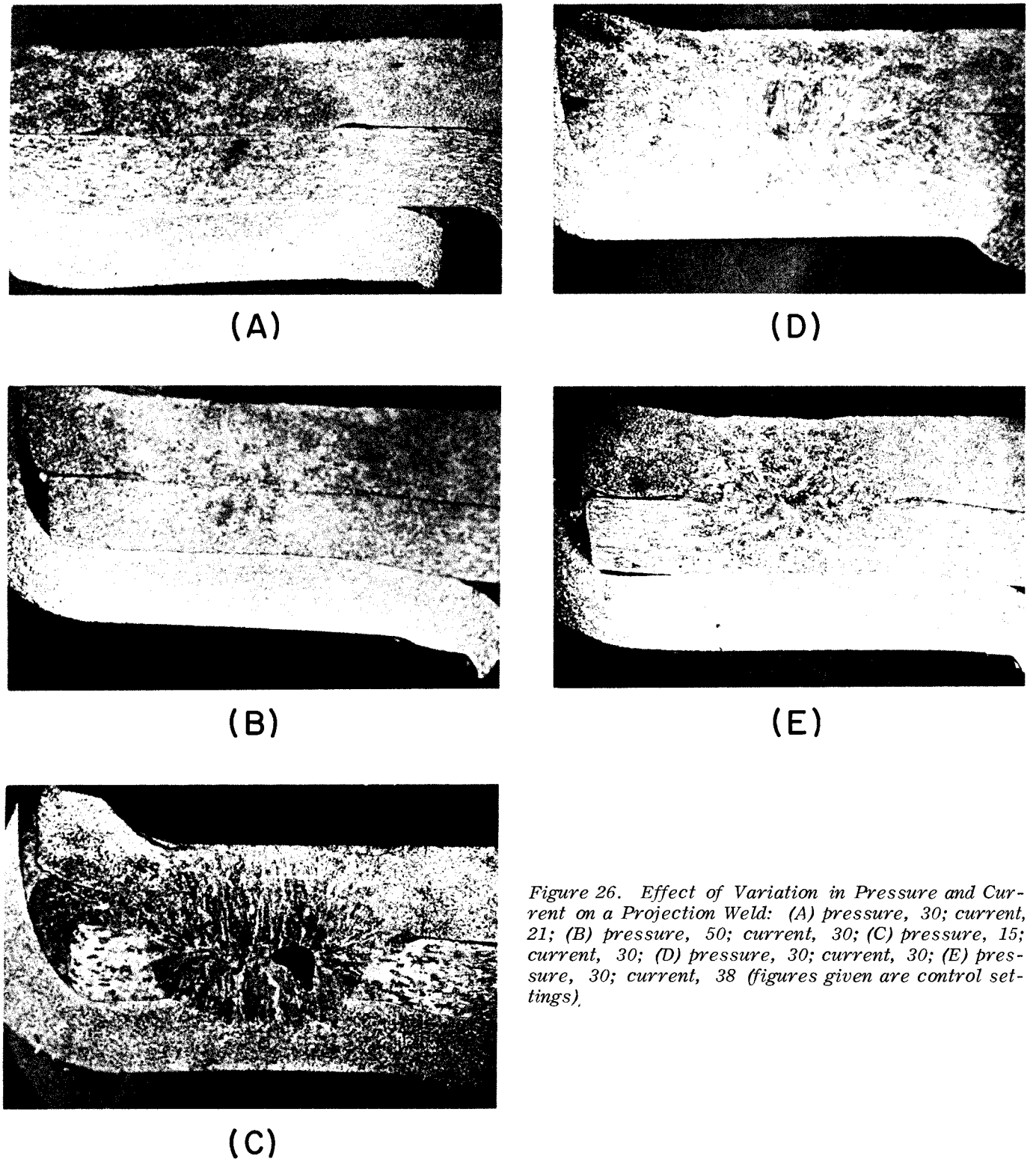


Figure 26. Effect of Variation in Pressure and Current on a Projection Weld: (A) pressure, 30; current, 21; (B) pressure, 50; current, 30; (C) pressure, 15; current, 30; (D) pressure, 30; current, 30; (E) pressure, 30; current, 38 (figures given are control settings).

tube it is used exclusively in pinching off the exhaust tubulation of annealed copper. After exhaust, the cold weld is made by pinching-off manually. A modified crimping tool is used which has for its jaws opposing 1/4-inch diameter hardened steel pins about 1 inch long. The operation may also be done by machine. The long axis of the tubulation is always kept normal to the long axis of the pins as the pinch-off is made. It has been shown that if the contacting, or inner, surfaces of the tubulation are clean and free of oxide, a true fusion weld in the metallurgical sense is obtained. Cold welds in tubulations of over 1/4-inch in diameter have been made successfully by this technique.

RIVETING, FOLDING, AND NOTCHING AND PEENING

Many of the metal-to-metal connections in tubes are simply mechanical connections where no attempt is made to fuse the parent metals. Riveting, folding, and notching and peening are all mechanical connections of this type.

The first of these, riveting, or staking as it was formerly called, is used almost exclusively for securing together the two halves of the plate. The operation (Fig. 27) consists of bending the two halves back-to-back (A) on the four-slide machine, and then piercing them as they are held together (B). The two flaps formed by the piercing operation at the point of entry of the piercing tool are then curled back on themselves by a die moving in a direction opposed to and from the side of the plate opposite the piercing tool (C). Several rivets may be made simultaneously. The stock used for riveting is usually mild cold-rolled steel between 5 and 15 mils thick.

A second type of joint used for plates and also, extensively, for cathode manufacture is known as a "lock seam" (Fig. 28). The process of making a lock-seam joint consists of folding the metal over on itself along one edge to enfold the other edge (A), and then folding both back on themselves (B). The stock used for cathodes is annealed nickel strip from 1.8 to 3.5 mils thick.

The final technique to be considered is notching and peening. It is mentioned here only for the sake of completeness, for it is considered more fully under the section dealing with grids and their construction. The operation is carried out completely on the grid lathe while the grid is being wound. Its purpose is to secure the grid lateral wires to the side rods. The side rods, being of a soft material are notched by a sharp notching roller. The lateral wires are laid in the notches and the notches are closed, or peened over, immediately. The depth of the notch and the hardness of the side rods are important variables in the quality of the connection. The "notch-peen" joint is found only on lathe-wound grids.

REFERENCE

1. Knoll, M. and B. Kazan, Materials and Processes for Electron Devices, p. 370, Springer Verlag, Berlin, Germany, 1959

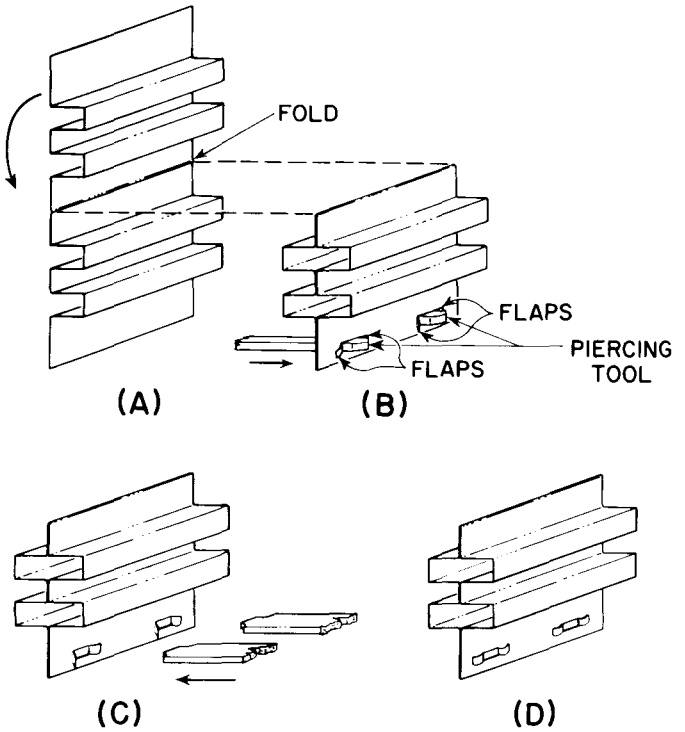


Figure 27. Riveting: (A) formed strip; (B) piercing folded strip; (C) curling and flattening flaps; (D) rivet completed

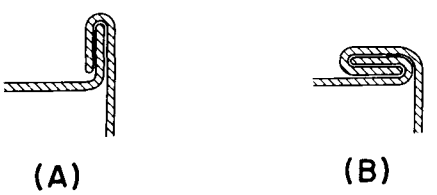


Figure 28. Lock Seam: (A) first fold; (B) completed lock seam

JOINING METALS TO GLASSES AND CERAMICS

K. Strater

The joining of metals to glasses and ceramic structures serves a number of important functions in the electron tube industry. These may be classified in the following categories:

Completion of electrical circuits
Support of tube parts
Completion of circuits and support for tube parts
Service as tube element
Window or screen as part of tube envelope

The reliability of these seals and of the techniques used to obtain them are reflected in their low incidence of failure during service. This is due to a seal-making technology which, in the case of metal-to-glass seals, is founded on well established principles. In the case of metal-to-ceramic seals, an empirical background has provided the principle basis for the present bonding techniques. However, recent work on the theoretical aspects of the adherence mechanism may lead to a broader application of metal-to-ceramic seals.

METAL-TO-GLASS SEALS

There are two general types of metal-to-glass seals: (1) metal-in-glass seals such as metal lead wires in a glass stem and top caps imbedded in a glass envelope; and (2) glass-in-metal seals such as the bonding of a metal rim to the perimeter of a glass stem.

The criterion for selecting a metal and a glass to be joined is that they have similar thermal expansion characteristics over the range of temperature to be used. However, in certain tube types, the use of thermally mismatched materials may be unavoidable. The primary factor that determines the leak-tightness and permanence of a metal-to-glass seal is the character and intensity of the internal stresses at the joint. Very often, the stress can be kept well below the fracture point, not only by selecting a glass and a metal of similar thermal expansion, but by designing the seal to take advantage of the high compressive strength of glass as compared to its tensile strength. Stresses may also be reduced by the use of a soft intermediate metal like copper which, by yielding, can dissipate stresses induced in the seal.

Fig. 29 gives an analysis of the stress inherent in a lead-wire seal when the metal has a higher coefficient of thermal expansion than the glass. If the glass and metal are of equal length when they are at the temperature at which they would be joined together, then, upon cooling, the glass, if unattached, would be longer than the metal. But, because they are joined together, the glass and metal are of equal length at room temperature, so that the glass has been compressed because of its adhesion to the metal and the metal has been placed in tension. This example is representative of the stress in the longitudinal direction. Similarly, a circumferential stress exists because the metal, firmly bonded to the glass, tries to pull it into a smaller compass. The resulting forces in the radial direction show that a radial tensile stress exists at the glass-to-metal boundary. Thus, the stresses resulting from the differential contraction between glass and metal can be divided into components in the longitudinal, radial, and circumferential directions. Table XVI indicates the relative stress distribution in this type of simple lead seal for the two conditions of thermal expansion characteristics:

Table XVI

Relationship Between Stress Distribution in Seals and the Relative Thermal Expansion of Metals and Glass

Thermal Expansion Coefficient	Stress in Glass		
	Longitudinal	Circumferential	Radial
Metal higher than glass	Compression	Compression	Tension
Metal lower than glass	Tension	Tension	Compression

Ideally, the thermal expansion curves of a metal and of a glass used in a seal should match over the full range of temperatures up to the setting point of the glass. If the expansion and contraction rates are different, the magnitude of the stresses that can be tolerated in the joint will depend primarily upon the tensile strength of the glass and the safety factor applied.

Experimental results reported by Partridge[1] suggest that a safety factor of four should be used in the design of seals which contain induced stresses. That is, the glass within the seal may be stressed up to one-quarter of its tensile strength. This factor is based upon the supposition that metal-to-glass seal failures are primarily a fatigue phenomenon whereby a repeated application of a stress, whether applied mechanically, by vibration, or thermally, by heating and cooling, causes the seal to fail. The correctness of this supposition would indicate that there is a theoretical maximum residual stress which the seal can sustain for an infinite time without cracking. In actual practice, it is impractical to determine this stress. However, it is general practice in the industry to limit the maximum permissible stress of a glass in a glass-to-metal seal to one kilogram per square millimeter.

If the seal is designed so that the resultant forces are compressive, the problem caused by thermal mismatch of the materials is considerably lessened because at room temperature the compressive strength of a glass is approximately ten times greater than the tensile strength. A rim seal joint in which a glass stem of lower coefficient of expansion is sealed to the inside of a metal envelope of higher coefficient of expansion is an example of a seal that, upon cooling, is compressively stressed.

Stress analysis of a seal is complicated by the fact that the strength of a piece of glass is not an absolute quantity. It depends more on its dimensions, shape, and surface finish than on its composition. The strength

of a seal will be affected by its contour; sharp re-entrant angles in the glass have a pronounced weakening effect. Further, the stress at room temperature gives no indication of the stress at service temperatures. Operating conditions must be considered in designing a satisfactory seal.

The glasses developed for bonding to metals fall into two general categories: (1) the soft glasses (soda-lime-silicate glasses) having a high coefficient of thermal expansion; and (2) the hard glasses (borosilicate type) which have a relatively low thermal coefficient of expansion. Their approximate ranges of expansion coefficients are:

(1) soft glass: 8.0×10^{-6} to 13.0×10^{-6} inches per inch per deg C
(2) hard glass: 3.0×10^{-6} to 6.0×10^{-6} inches per inch per deg C

The hard glasses have higher annealing temperatures and can, therefore, be used at higher operating temperatures. An extensive treatment of the properties of specific glasses is included in the chapter on Glass for Receiving Tubes by J. Gallup.

The choice of pure metals of similar coefficients of expansion to the glasses is limited to platinum, tungsten, and molybdenum, but the development of iron, nickel, chromium, and cobalt alloys has widened the range of available metals (Table XVII). The thermal expansion curves of some of these metals and the two types of glass are shown in Fig. 30.

The advantages of platinum as a metal for use in glass-to-metal joints are:

1. Its thermal expansion coefficient is similar to that of many soft glasses so that there is no difficulty in obtaining a close match.

2. It possesses a very low yield-strength and is ductile; thus it can relieve the strain in the glass component of seal.

3. It is resistant to oxidation; only a thin adherent oxide film is formed.

Although platinum has many advantages, its cost precludes its use in large-scale manufacture.

Dumet has been developed as a substitute for platinum. It consists of a central core of a 43 per cent nickel alloy coated with about 20 per cent of its weight of electrodeposited or clad copper. The composite rod is swaged and drawn into a wire having a radial coefficient of expansion similar to that of platinum (Table XVII), and a longitudinal coefficient of expansion closer to that of the lower-expansion nickel alloys. The radial expansion is controlled by the ratio of copper sheathing to core material, and increases with increasing thickness of copper.

Dumet alloy is used in the mass-production of lead-wire seals imbedded in glass stems. Such seals are not fully annealed so that the glass is left in compression in the longitudinal direction; in the fully-annealed condition

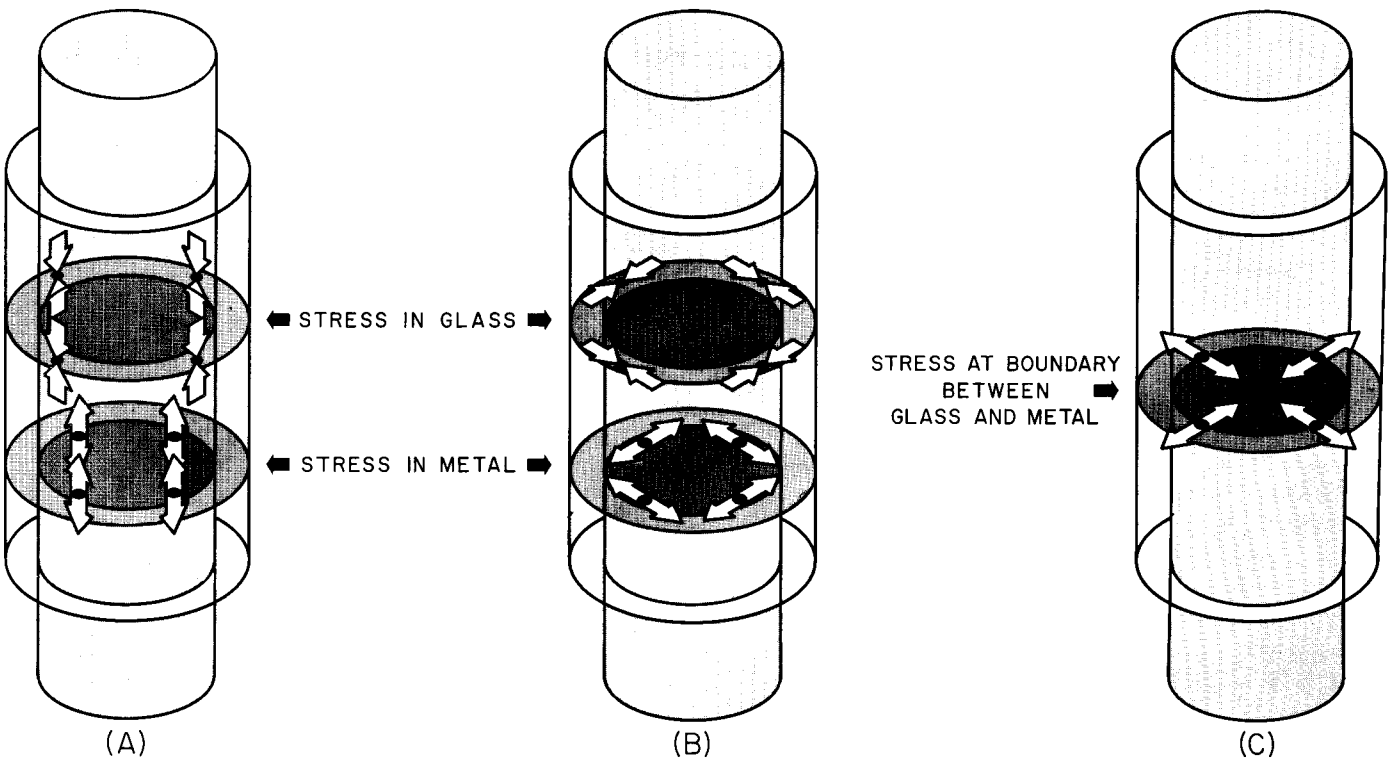


Figure 29. Stresses at Room Temperature in a Lead Seal Due to Differential Contraction of Metal Lead Wire and Glass Stem (coefficient of thermal expansion of metal higher than that of glass): (A) longitudinal stresses; (B) tangential stresses; (C) radial stresses

the glass would be in tension. To further reduce the build-up of dangerous stresses, small-diameter lead wires, usually not exceeding 0.020 inch are used. Although the dumet section of lead wires often extends through the glass seal, the length of the wire used is usually restricted to a length within the bounds of the seal. The connecting wires, generally of a larger diameter, are imbedded in the ends of the seal.

Tungsten, molybdenum, and their alloys are used for sealing to the low-expansion glasses. They are hard and brittle metals which are difficult to work so that care must be taken to avoid splitting of the metal both in fabricating it into rod and wire, and in cutting the finished material into the short lengths needed for seal making. In addition, these materials oxidize readily at comparatively low temperatures. Thin oxide coatings on tungsten wire are fairly adherent, but they can be easily disturbed while the wire is being inserted in the glass walls of an electrode assembly. Molybdenum rapidly becomes covered with a thick, nonadherent film of oxide so that contact with air must be restricted while the molybdenum is being coated with glass.

Over a limited range of temperature, certain nickel-iron alloys also have expansion properties similar to those of the hard glasses. However, expansion coefficients increase above temperatures corresponding to the loss of ferromagnetism. For example, the linear expansion coefficient of a 36 per cent nickel-iron alloy is 2.6×10^{-6} inches per inch per degree C over the temperature range 0 to 200 C, but it is as high as 6.9×10^{-6} inches per inch per degree C between 200 C and 250 C. The inflection point of this alloy, or the point at which the slope of the coefficient of thermal expansion versus temperature suddenly increases, is approximately 150 C. Increasing the nickel content of the alloy has a marked effect upon both the coefficient of expansion and the inflection temperature. The thermal expansion curve of a 50 per cent nickel-iron alloy (52 metal) is practically a straight line and shows no inflection over a temperature range of 0 to 500 C (Fig. 30). The actual inflection temperature has been raised to 515 C. In addition, the increased nickel content has raised the average thermal coefficient of expansion to 9.18×10^{-6} inches per inch per degree C. From these data, it can be seen that low-expansion nickel-iron alloys are not suitable for sealing to hard glasses because the inflection temperature is so far removed from the transition temperature of the glass that joints would crack upon cooling. For satisfactory seals, the inflection temperature of the alloy should not be lower than the annealing range of the glass. This consideration limits the choice of a nickel-iron alloy to the higher-expansion alloys containing at least 46 per cent nickel.

Care must be taken to avoid over-oxidation of these alloys because their oxide films are not firmly adherent to the metal. Therefore, in order to keep the thickness of the oxide film to a minimum during sealing, only the reducing portion of the flame should be used.

Partial substitution of cobalt for nickel in the nickel-iron alloys lowers the thermal expansion coefficient without affecting the inflection temperature and results in alloys suitable for sealing to hard glasses. This low-

Table XVII

Common Metals Used in Metal-To-Glass Seals

Metal	Nominal Coefficient of Linear Thermal Expansion Over the Range 20 to 350 C — inches per inch per deg C	RCA Specification Number
Platinum	9.4×10^{-6}	
Dumet (composite alloy: Sheath - 100% copper; Core - 42% nickel, 58% iron alloy; total composition, 20% copper, 80% nickel-iron alloy)	9.0×10^{-6} (radial) 7.1×10^{-6} (longitudinal)	D-7
27% Chromium-Iron Alloy (27% chromium, 73% iron)	10.2×10^{-6}	S-73
No. 4 Alloy (6% chromium, 42% nickel, 52% iron)	8.9×10^{-6}	C-180
52 Metal (50% nickel, 50% iron)	9.5×10^{-6}	N-71
Cold-Rolled SAE 1010 Steel	13.7×10^{-6}	S-10
Kovar (17% cobalt, 29% nickel, 54% iron)	4.7×10^{-6}	I-50
Molybdenum	5.5×10^{-6}	M-13
Tungsten	4.4×10^{-6}	T-10

ering of expansion increases as the proportion of nickel is replaced by cobalt to a limit of about 18 per cent. A comparison of the thermal expansion coefficients for Kovar (17 per cent cobalt, 29 per cent nickel and 54 per cent iron) and 52 alloy (50 per cent nickel and 50 per cent iron) (Fig. 30) reflects this change in expansion characteristics.

Cobalt also lowers the electrical resistivity of nickel-iron alloys, thus reducing the heating effects in the seal when current flows. Cobalt also results in a more fusible oxide when the metal is heated in readiness for coating with glass.

In contrast to cobalt, chromium is added to nickel-iron alloys to raise their thermal expansion coefficients to the level of the soft glasses. A range of expansions can be obtained by substituting chromium for iron in the 42 per cent nickel-iron base alloy, and minor adjustments can be made to the expansion by altering slightly the proportions of nickel and chromium. The expansion curve for No. 4 alloy (6 per cent chromium, 42 per cent nickel and 52 per cent iron) (Fig. 30), shows this increase.

Alloys containing chromium and iron have thermal coefficients which match those of the soft glasses. These alloys are probably the most resistant to oxidation of all the metals used for sealing to glass, with the exception of the noble metals. The thin oxide film is rich in chromium oxide and adheres tenaciously to the base metal, which it protects from further oxidation during the seal-making process.

In spite of its thermal expansion of 16.7×10^{-6} inches per inch per degree C, copper can be sealed to a wide range of glasses if a suitable form of seal is used. Its low strength and high ductility permit deformation and consequent relief of stress in the glass. The technique developed for sealing utilizes the Housekeeper principle of sealing the glass to a feathered edge of copper not exceeding 0.010 inches. This technique allows the copper to yield well below the rupture strength of the glass.

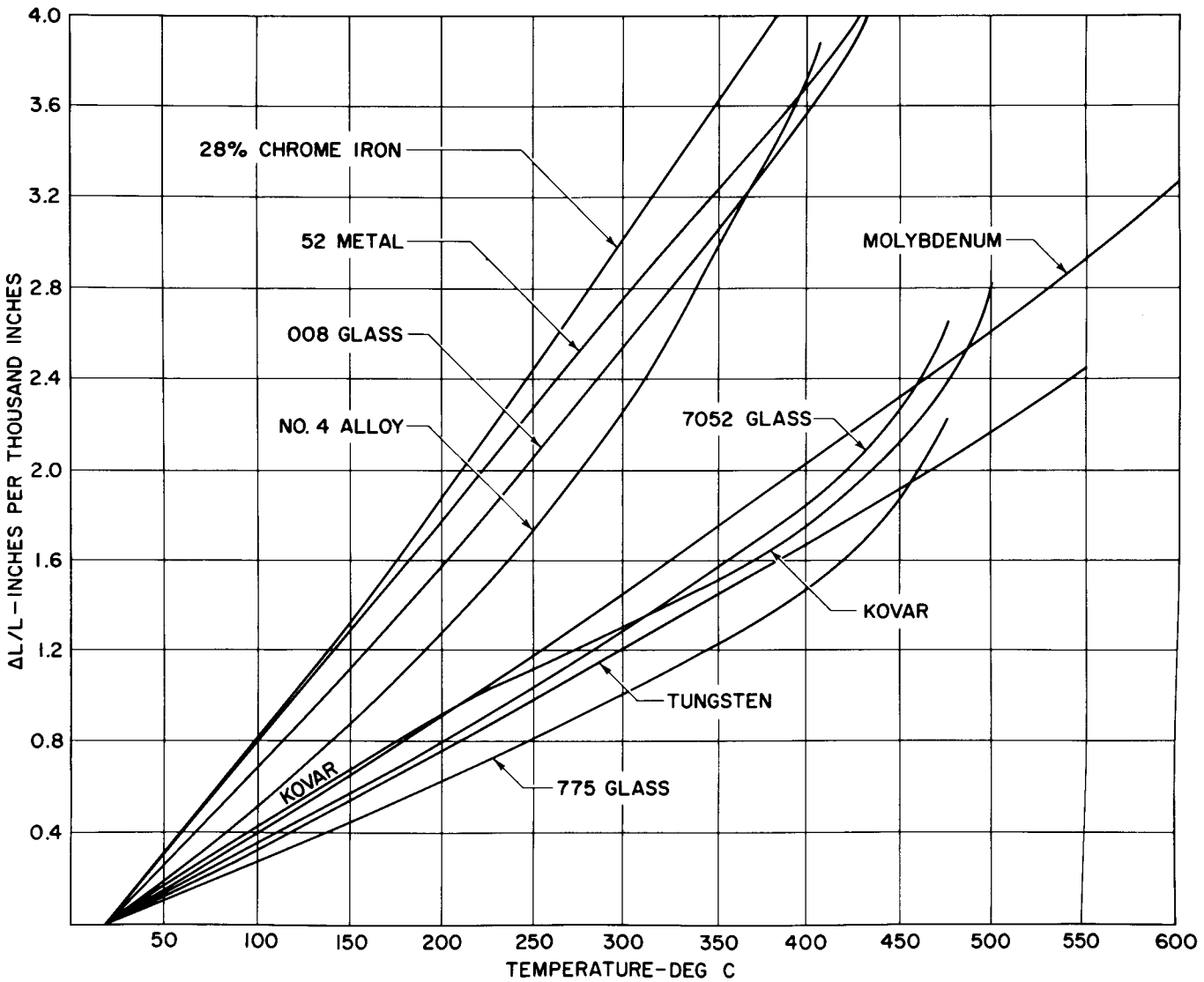


Figure 30. Thermal Expansion of Some Metals and Glasses Used in Metal-to-Glass Seals

The relationship between the expansion curves of various metals and glasses (Fig. 30) indicates that the magnitude of the stress incurred is affected by the choice of temperature range over which the thermal expansions are compared. Whereas metals, with the exception of those having inflection points, possess a nearly straight-line expansion curve, the expansion of glass increases rapidly before the setting point is reached. The transition point for glass, or the point at which the slope of the coefficient of thermal expansion versus temperature suddenly increases, occurs at a temperature lower than the lowest annealing temperature, so that it is impossible to match the glass expansion over the range of temperature from the setting point to room temperature with the nearly straight-line expansion of pure metals. For many seals, the best compromise is to find combinations of glass and metal that match at the setting point and at room temperature. As illustrated in Fig. 29, the final strain within the seal is the result of the difference in contraction which would exist between the glass and the metal if they were free to contract separately from slightly below the setting point to room temperature.

As previously indicated, a major factor in producing an adequate metal-to-glass seal is the formation of a thin, adherent oxide film on the metal in contact with the glass. Some of the metal oxide dissolves in the glass at the sealing temperature and thus provides a bonding layer between the metal and the glass.

Because some metals oxidize very readily, the supply of oxygen can be limited by either coating the metal with a thin film of fused borax, or by using a nitrogen atmosphere during the heating step preparatory to fusing with glass. Oxidizing conditions during actual sealing can also be suppressed by using the reducing portion of the sealing flame.

In its initial stage, the rate of oxidation of the metal depends on the affinity of the metal for oxygen; however, the first continuous film formed in the initial stage serves as a barrier to prevent direct contact of oxygen with the metal. Further oxidation then proceeds by the movement of ions through this oxide film, which consists of a lattice of electropositive metal ions and electronegative oxygen ions. The metal ions leave the metal, enter the oxide, and move outward towards the oxide-air boundary where they come in contact with the free oxygen. Since the formation of the initial film is dependent on the chemical affinity of the metal for oxygen, alloy steels such as the chromium-iron alloys oxidize more rapidly at first than carbon steel. However, they form adherent protective scales which retard further oxidation, and thus, prevent the formation of excessively thick oxide films.

METAL-TO-CERAMIC SEALS

The use of ceramics in place of the glass components in receiving tubes has the advantage of permitting higher operating temperatures in addition to allowing higher bake-out temperatures, thus reducing the amount of residual gases in the tube. Since residual gases eventually lead to tube failure, ceramic or combined ceramic-and-metal tubes should be more reliable than the present glass tubes.

One of the primary obstacles to producing ceramic tubes is the difficulty in obtaining a mechanically strong, leak-tight bond between materials as inherently dissimilar as metals and ceramics.

The two fundamental methods of bonding a metal member to a ceramic body or of bonding two ceramic bodies are: (1) forming a thin, tightly adherent metallic layer on the surface of the ceramic, thus providing a surface which can be wet by a brazing alloy and, in this manner, easily bonded to a metal or another ceramic; and (2) utilizing a brazing alloy (containing an active metal such as titanium or zirconium) which reacts with, and wets, the ceramic surface. In both cases metal-to-ceramic adherence is promoted by a strong chemical reaction. The thin metallic layer obtained by the first method is usually molybdenum which is bonded to the ceramic under closely controlled conditions of temperature and atmosphere. Current experimental evidence indicates that a reaction probably occurs between an oxide of the coating metal and a low-melting point phase within the ceramic body. In the active metal process, the reaction between the titanium or zirconium constituent of the brazing alloy can only occur under vacuum. At an appropriate temperature and vacuum the active metallic element has a greater affinity for oxygen than some of the constituents in the ceramic structure and, consequently, an exchange of bonding occurs in which the active element becomes as tightly bonded to the oxides in the ceramic as it is alloyed with the primary metal of the brazing alloy.

The most widely used bonding method is the molybdenum coating process. The coating can be applied either as a powder on the surface of the ceramic or by dipping the ceramic into a concentrated solution of a soluble molybdenum compound. Powered molybdenum, suspended in an organic binder, with or without additional elements such as iron, cobalt, nickel or manganese, can be applied to a ceramic surface by painting, spraying or silk screening.

The silk-screening method[2] for a specific alumina ceramic has been developed to a high degree of uniformity for a large-scale operation. The dimensions of the unfired coating is controlled to a thickness of 0.0015 ± 0.0003 inch which allows for a 30 per cent shrinkage during firing. This coating thickness uniformity is effected by rigid control of the molybdenum particle size to particles of one micron and smaller. The coated ceramic, containing molybdenum particles suspended in ethyl cellulose and Dibutyl Cellosolve, is then dried in air and fired in a reducing atmosphere of saturated 10 per cent forming gas at 1450 C for one-half to one hour. The resultant tightly bonded coating appears gray. After firing, the ceramic is ready to be brazed to a metal or to another metallized ceramic.

The limitations on the thermal expansion of the ceramic body and metal components used in making a ceramic-to-metal seal are not as stringent as those required for a metal-to-glass seal because an intermediate copper brazing alloy is used which is capable of yielding and thus reducing the stress. In addition, the ceramic body is much stronger than glass. However, an excessively wide divergence in the thermal expansion

properties of a metal-ceramic combination would probably result in failure of the joint, especially if a not-too-ductile brazing alloy is used.

Proper design can do much to neutralize divergencies in expansion properties. An example is the nuvistor triode. It is constructed with a forsterite ceramic stem, containing molybdenum lead wires, and an envelope of 52 alloy (50 per cent nickel, 50 per cent iron) which is joined to the rim of the stem disc. The thermal expansion of the ceramic stem is less than that of the nickel-alloy envelope so that upon cooling, the envelope grips the stem, whereas the stem, having a higher thermal expansion than the molybdenum wire, contracts around the lead wire. Thus, both joints are in radial compression. A comparison of thermal expansion curves of forsterite and of alumina (Fig. 31) with those of several metals illustrates this difference.

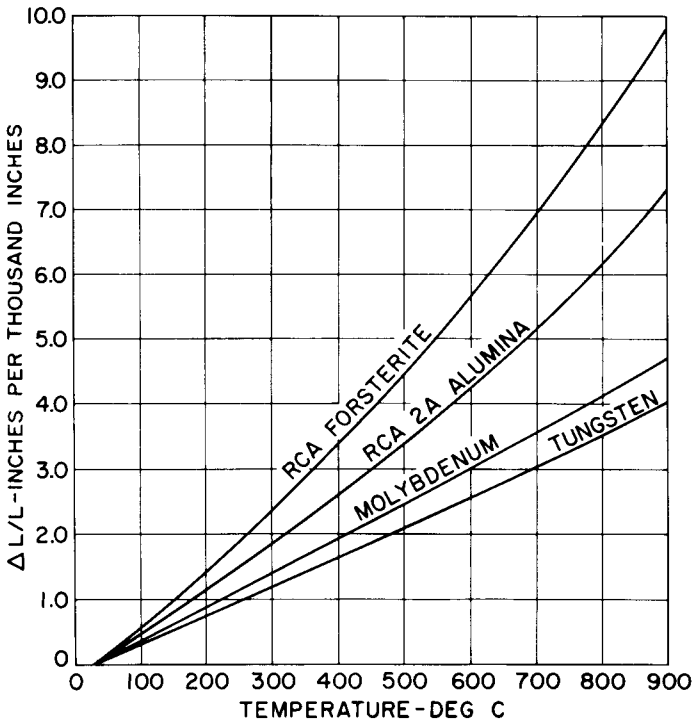


Figure 31. Thermal Expansion of Some Metals and Ceramics Used in Metal-to-Ceramic Seals

The weakest member in this type of joint is the ceramic body, which by its refractory nature, is brittle and, therefore, has low tensile strength, but relatively high compressive strength. For this reason, the metal-to-ceramic joint described above has been designed to use the most favorable strength property of the ceramic. As a general rule, the tensile strength of the ceramic represents a major limiting factor in the design of a tube. A bond which tends to place the ceramic in radial tension reduces the available tensile forces that may be applied to the tube during service.

Other important properties of a ceramic for tube application are low porosity, or vacuum tightness, good thermal conductivity (to dissipate heat readily), and high thermal-shock resistance. These properties are primarily functions of the raw materials and the processing methods which are used in their manufacture. A detailed discussion correlating these properties with structure and chemical composition is contained in the section on Ceramics by W. Lawrence.

Of the two molybdenum metallizing techniques, the more versatile is the solution method[3] in which a highly concentrated solution of a soluble molybdate, such as lithium dimolybdate, is used to cover the ceramic by a simple dipping process. Lead holes and other recessed areas, inaccessible by the powder technique, can readily be coated by this method. The ceramic part to be metallized is first solvent cleaned and is then dipped into the metallizing solution. Excess solution is shaken off and the part is allowed to dry in air.

If the ceramic is the RCA developed forsterite body, it is then fired in air at 700 C for 15 minutes during which the lithium dimolybdate melts and flows over the surface of the ceramic stem. The excess metallizer drains away, resulting in a reproducibly uniform thickness of coating. This uniformity is important in lead holes where thickness tolerances are critical.

The next step in the process is the reduction of the lithium dimolybdate to molybdenum metal in a reducing atmosphere of "wet" hydrogen. The reducing conditions,[2] temperature, atmosphere, time at temperature, and the heating rate[2] are dependent upon characteristics of the ceramic body, such as the relative amounts of the various oxide phases, their chemical composition and melting points, the softening temperature of the body, and its thermal-shock resistance.

In the heating of a ceramic body, if a large difference exists between the skin temperature and the core temperature, different rates of thermal expansion will be induced, and internal stresses may build up to the point where they relieve themselves by forming a crack. Therefore, the size of a ceramic body will have an appreciable effect upon the rate at which a coated part may be introduced into the reduction furnace.

The optimum reduction treatment for forsterite stem discs was found to be 1300 C for approximately 15 minutes. As for the RCA alumina body, the optimum reduction temperature is somewhat higher, approximately 1400 to 1500 C. This difference results from the higher melting point of the constituent phases within the alumina body which necessitates higher temperature to ensure that sufficient reaction occurs.

An important factor in the reduction process, one that has only recently been appreciated, is the composition and moisture content of the reducing atmosphere. Recent work on molybdenum metallizing of alumina[3] has established a relationship between the water-vapor concentration and metal-to-ceramic bond strength over a range of reducing atmospheres from 30 per cent forming gas to pure hydrogen. The bond strength improves with increasing amounts of water vapor up to a maximum, beyond which further increase in water vapor has no effect. The bond strength was found to vary with the reducing atmosphere expressed as a ratio of the partial pressures of water vapor to hydrogen, thus permitting an interchange of reducing atmospheres (various hydrogen and

nitrogen mixtures) at calculated moisture concentrations sufficient to obtain good adherence. It is postulated that the water vapor in the reducing atmosphere aids in the retention of molybdenum dioxide up to a temperature at which the oxide reacts with the low-melting-point constituents of the ceramic body, thus forming a more adherent bond.

The second fundamental type of metallizing process, the active-metal process,[4] utilizes a copper or nickel brazing alloy containing titanium or zirconium. When the brazing alloy is molten, the active metal reacts with some of the oxide constituents in the ceramic body forming a complex titanium or zirconium oxide which is still bonded to the copper in the brazing alloy.

A typical brazing alloy can be made by mixing powders of nickel or copper with titanium to form alloys of 65 per cent nickel -35 per cent titanium and of 65 per cent copper -35 per cent titanium. These powders are then pressed into washers, rings, or other shapes convenient for brazing. A more advanced technique utilizes a process whereby brazing washers are pressed directly onto the lead wires. Experimental results in the formation of the optimum alloy compositions indicated that alloys having a higher titanium concentration, such as those discussed above, wet and flowed on the ceramic more easily than those with lesser concentrations.

Molybdenum lead wires can be brazed to an alumina stem in a vacuum induction furnace under a vacuum of approximately 0.2 micron. The parts to be brazed are assembled on a jig, which is placed on a movable platform inside the induction coil. The induction coil has a molybdenum susceptor on the inside so that both the metal and the ceramic parts are heated uniformly by radiation.

If the nickel alloy is being used for stem brazing, the temperature of the stem assembly is brought up to 1250 C in four minutes; if the brazing medium is the copper alloy, the temperature is increased to 1050 C in three minutes. The heating time is specifically defined since the quality of the fillet is dependent upon it. When the melting point of the brazing alloy is approached slowly, the titanium oxidizes even at the low pressure and drastically reduces the wetting characteristics of the brazing alloy, so that little or no flow of the alloy occurs. The rate of heating to 800 C is unimportant since titanium begins to absorb gases only at higher temperatures. From 800 C to the melting point, the rate of heating should be as rapid as possible.

After brazing, the metal-ceramic assembly is cooled in vacuum to 500 C in ten minutes. At this point, helium or hydrogen is added to increase the cooling rate to 200 C. The furnace chamber is then opened and the brazed assemblies are removed.

Other techniques which utilize oxide powders, or metal powders other than molybdenum, can be used to obtain a coating on a ceramic.[5] Similarly, an active brazing alloy can be made by using hydrides of titanium or zirconium. The methods outlined above, however, represent the most promising methods to date of joining a metal to a ceramic.

REFERENCES

1. Partridge, J.H., Glass-to-Metal Seals, Society of Glass Technology, 1949
2. Code L2
3. Code L3
4. Stoeckert, A. J., "A Novel Method of Fabricating Ceramic Stems," Proceedings 1959 Electronic Components Conference, May 1959
5. Code L4

SYSTEMATIC CHARTS OF THE PROPERTIES OF ELECTRON TUBE METALS

S. Y. Husni

This section contains tabular and graphic information on some of the properties of metals and alloys that should be of interest to the electron tube engineer. The following tables and charts are included:

Properties of Metals . . . . . Table XVIII

Properties of Alloys . . . . . Table XIX

Coefficients of Thermal Expansion at Elevated Temperatures . . . . . Table XX

Tensile Strengths at Elevated Temperatures . . . . . Table XXI

Total Emissivity . . . . . Table XXII

Thermal Conductivity at Elevated Temperatures . . . . . Table XXIII

Charts of Vapor Pressures at Elevated Temperatures . . . . . Figs. 32, 33

Table XVIII

Properties of Metals

Metal	Density g/cc (at 20 deg C)	Melting Point deg C	Temperature for 10^{-5} atm Vapor Pressure— deg C	Thermal Conductivity watts/cm²/cm/ deg C/sec (near 20 C)	Thermal Conductivity cal/cm²/cm/ deg C/sec (near 20 C)	Thermal Expansion microinches/ deg C (at 20 C)	Specific Heat— cal/g/deg C (at 20 C)
Aluminum 2S A31*	2.70	660	1190	2.22	0.53	23.9 (20-100 C)	0.215
Barium	3.5	704 ± 20	615			18	0.068
Chromium	7.19	1890 ± 10	1380	0.67	0.16	8.2 (0-100 C)	0.11
Cobalt	8.90	1495	1500	0.71	0.17	12.3	0.099
Columbium+	8.57	2415 ± 15		0.54	0.13	7.1	0.065
Copper	8.89	1083	1280	3.93	0.94	16.5	0.092
Gold	19.3	1063	1320	2.97	0.71	14.2	0.031
Hafnium	11.4	2110		0.21	0.05	6.2	0.035
Iridium	22.5	2454	2280	0.59	0.14	6.8	0.031
Iron	7.87	1539	1430	0.75	0.18	11.7	0.11
Lead	11.3	327		0.34	0.08	29.3 (20-100 C)	0.031
Magnesium	1.74	650	425	1.59	0.38	26 (40 C)	0.25
Molybdenum	10.2	2625 ± 50	2630	1.59	0.38	5.3 (0-20 C)	0.062
Nickel	8.9	1455	1470	0.67	0.16	13.3 (0-100 C)	0.105
Osmium	22.5	2700 ± 200				4.6	0.031
Palladium P39*	12.0	1554	1500	0.71	0.17	11.8	0.058 (0 C)
Platinum P12*	21.5	1774	2050	0.71	0.17	8.9	0.032
Rhenium	20.5	3170 ± 60	3000	0.71	0.17	4.7 to 12.5	0.033
Rhodium	12.4	1966	2025	0.88	0.21	8.3	0.059
Ruthenium	12.2	2500 ± 100				9.1	0.057 (0 C)
Silver	10.5	961	1030	4.18	1.0	19.7 (0-100 C)	0.056 (0 C)
Tantalum T13*	16.6	2996 ± 50	3020	0.54	0.13	6.5	0.036 (0 C)
Tin	7.30	232		0.67	0.16	23	0.054
Titanium T6*	4.54	1820 ± 100	1520	1.72	0.41	8.5	0.126
Tungsten T10*	19.3	3410 ± 20	3280	2.01	0.48	4.3	0.032
Thorium	11.5	1800 ± 150		0.37	0.088	11.1 (20-60 C)	0.034
Vanadium	6.0	1735 ± 50	1820	0.31	0.074	7.8	0.120
Zirconium	6.5	1875 ± 25	2400	0.16	0.040	5.9	0.068

*RCA Code.

+Niobium

Table XVIII

Properties of Metals

Tensile Strength psi	Yield Strength psi	Modulus of Elasticity psi	Per Cent Maximum Elongation	Electrical Resistivity microhm-cm (at 20 C)	Temperature Coefficient of Resistivity	Present or Potential Use in Electron Tubes
13×10^3	5×10^3	40×10^6	45	2.655	44×10^{-4}	surface cladding for plates and getters
				50-75		getters
		36×10^6		13 (28 C)		metallizing
37×10^3	34×10^3	30×10^6	8	6.24	33×10^{-4}	metallizing
50×10^3		12×10^6	49	13.1 (18 C)	39.5×10^{-4}	getters, heater wires
32×10^3		16×10^6	45	1.673	68×10^{-4}	lead wires and support cups; surface cladding for lead wires; surface plating for lead wires, grid siderods, grid lateral wires, and grid frames; brazing lead wires to plates, grid lateral wires to grid frames, and grid lateral wires to grid siderods
19×10^3		12×10^6	77	2.19 (0 C)	34×10^{-4}	surface plating grid lateral wires
68×10^3	34×10^3	20×10^6	21	32.4	44×10^{-4}	heater wires
36×10^3		75×10^6		5.3	40×10^{-4}	heater wires
45×10^3		29×10^6	40	9.71	62×10^{-4}	metallizing, grid siderods, grid lateral wires, lead wires
2×10^3	1×10^3	3×10^6	47	20.6	33.6×10^{-4} (20-40 C)	solder
32×10^3	25×10^3	7×10^6		4.46	40×10^{-4}	getter
220×10^3 63×10^3†	56×10^3	50×10^6	30	5.17 (0 C)	47×10^{-4}	lead wires, grid lateral wires, grid siderods, grid frames, heater wires, heater spacer wires, metallizing, plates
29×10^3 8×10^3†	9×10^3	30×10^6	50	6.84	48×10^{-4}	metallizing, plates; surface plating on top caps, base pins, lead wires, headers and insert, plates, grid lateral wires, grid siderods, grid frames, support cups, and metallizing; surface cladding on lead wires, plates, and grid siderods
		80×10^6		9.5	42×10^{-4}	heater wires
24×10^3	5×10^3	17×10^6	40	10.8	37×10^{-4}	surface plating on grid lateral wires
18×10^3	2-5.5 $\times 10^3$	21×10^6	40	9.83 (0 C)	30×10^{-4}	plates, cathodes, surface plating on grid lateral wires
				21	31×10^{-4}	heater wires
80×10^3		42×10^6		4.5	46×10^{-4}	surface plating on grid lateral wires
		60×10^6		7.6 (0 C)		heater wires
18×10^3	8×10^3	11×10^6	50	1.59 (0 C)	38×10^{-4}	surface plating on base pins, lead wires, header and inserts, and grid siderods
50×10^3		27×10^6	40	12.4 (18 C)	38.2×10^{-4}	lead wires
2×10^3	1×10^3	6×10^6	47	20.6	33.6×10^{-4}	solder
95×10^3	81×10^3	17×10^6	22	80 (0 C)	54×10^{-4}	getters, metallizing
300×10^3 130×10^3†	44×10^3	50×10^6	10	5.5	48×10^{-4}	grid lateral wires, heater wires, heater spacer wires, lead wires, metallizing
38×10^3	27×10^3	10×10^6	40	19		getters
72×10^3	64×10^3	20×10^6	28	26		
50×10^3 34×10^3†	30×10^3	14×10^6	26	39.7	44×10^{-4}	getters, metallizing

†Measured at 800 C.

Table XIX

Properties of Alloys

Alloy•	Density g/cc (at 20 deg C)	Melting Point deg C	Thermal Conductivity watts/cm²/cm/deg C/sec (near 20 C)	Thermal Conductivity cal/cm²/cm/deg C/sec (near 20 C)	Thermal Expansion microinches/deg C (at 20 C)	Specific Heat— cal/g/deg C (at 20 C)	Tensile Strength psi
Brass (Cartridge)	8.53	955	1.21	0.29	19.9 (25-300 C)	0.09	44×10^3
Carpenter 426[13]	8.12	1460	0.12	0.029	11.9 (20-600 C)		70×10^3
Chromiron C72*	7.6	1470 ± 50	0.21	0.05 (100 C)	10.8 (0-100 C)	0.44 (100 C)	80×10^3
Constantan	8.9	1210	0.23	0.054	17	0.098	
Copper Beryllium	8.23	955	0.84	0.20	16.6 (20-100 C)	0.1 (30-100 C)	72×10^3
Copper (Chrome) C92*	8.92	1083					37×10^3
Copper (Phosphorus Deoxidized)	8.94	1083	3.38	0.81	17.7 (20-300 C)	0.092	32×10^3
Copper (OFHC) C600*	8.92	1083	3.89	0.93	17.7	0.092	32×10^3 12×10^3 †
Copper (Tough-Pitch)	8.94	1083	3.89	0.93	17.7	0.092	32×10^3
Cupro Nickel	8.9	1240	0.29	0.07	16.2 (20-800 C)	0.09	55×10^3
Duranickel[14]	8.3	1440	0.63	0.15	13	0.13	105×10^3 20×10^3 †
Hastelloy A[17] H1*	8.8 (25 C)	1330	0.17	0.04	15.4	0.094	115×10^3 47×10^3 †
Hastelloy B[17] H26*	9.2	1333	0.11	0.027	14.6	0.091	130×10^3 70×10^3 †
Inconel[4]	8.47	1410 ± 15	0.15	0.036	11.5	0.109	85×10^3 32×10^3 †
Inconel X[4]	8.30	1410 ± 15	0.15	0.036	13.7	0.105	115×10^3 46×10^3 †
Invar[4] S160*	8.1	1425	0.11	0.026	1.0 (20-100 C)	0.123	70×10^3 20×10^3 †
Karma[1]	8.1	1400	0.13	0.031	13.3 (20-100 C)	0.104	130×10^3
Kovar "A"[6]	8.4	1450	0.19	0.046	5.1 (20-100 C)		75×10^3
Molybdenum-Tungsten D18*	13.4						185×10^3
Molybdenum-Tungsten H3W*	11.2						175×10^3
Molybdenum-Tungsten M8*	11.5						180×10^3
Monel[4]	8.8	1325 ± 25	0.25	0.06	14 (25-100 C)	0.127	75×10^3 16×10^3 †
Monel "K"[4]	8.5	1335 ± 25	0.18	0.043	14 (25-100 C)	0.127	100×10^3 25×10^3 †
Nichrome[1]	8.3	1350	0.13	.032	17 (20-1000 C)	0.107	105×10^3
Nichrome V[1] N23*	8.4	1400	0.13	.032	17.6 (70-1000 C)	0.107	95×10^3
Nickel "A"[4] N5*	8.9	1440	0.59	0.14	13.3 (25-100 C)	0.13	72×10^3 24×10^3 †
Nickel "D"[4] N13*	8.8	1435	0.42	0.10	14.3 (20-500 C)	0.127	87×10^3 20×10^3 †
Nickel "E"[4] N40*	8.7	1435	0.50	0.12	14.6 (20-500 C)	0.127	75×10^3
Nickel "N1W"[14] N1W*	8.77	1457					83×10^3 29×10^3 †
Nickel "N91"[14] N91*	8.98	1443					18×10^3 †
Nickel "N97"[14] N97*	8.88	1449					18×10^3 †

Table XIX

Properties of Alloys

Yield Strength psi	Modulus of Elasticity psi	Per Cent Maximum Elongation	Electrical Resistivity microhm-cm (at 20 C)	Temperature Coefficient of Resistivity	Present or Potential Use in Electron Tubes	Approximate Composition (per cent)	Other Designations•
11×10^3	16×10^6	66	6.2	14.8×10^{-4}	top caps, base pins	73 Cu 30 Zn	
30×10^3		35	94	25×10^{-4}	top caps	42 Ni 6 Cr Bal. Fe	Sealmet 4[3]
50×10^3	29×10^6	25			headers and insert	25 Cr 1.5 Mn 1 Si Bal. Fe	446 Stainless Almet 446[5]
			44.1	0.08×10^{-4}		40 Ni 60 Cu	Advance[1] Copel[11] Eureka[16]
25×10^3	17×10^6	35	10		grid siderods	97 Cu 2.2 Be	
		35	2.3		grid siderods	99 Cu 1 Cr	
10×10^3	17×10^6	45	2.03	30×10^{-4} (20-200 C)	lead wires	99.9 Cu 0.02 P	
10×10^3	17×10^6	50	1.76		lead wires, grid siderods	100 Cu	
10×10^3	17×10^6	45	1.71	39.2×10^{-4}	lead wires	99.9 Cu 0.04 O	
20×10^3	22×10^6	45	37		getter supports	30 Ni 70 Cu	
50×10^3	30×10^6	35	45	36×10^{-4}	grid lateral wires	93 Ni 5 Al 1 Si	Nickel "Z" Type A
49×10^3	27×10^6	45	126.7		lead wires, grid lateral wires	53 Ni 22 Mo 22 Fe 1 Si 2 Mn	
60×10^3	31×10^6	40	135		lead wires, grid lateral wires	60 Ni 32 Mo 6 Fe 1 Si 1 Mn	
35×10^3	31×10^6	45	98.2			76 Ni 16 Cr 7 Fe	
50×10^3	31×10^6	50	122.0			73 Ni 15 Cr 2.5 Ti 1 Cb 7 Fe	
		40	85	13.5×10^{-4} (20-100 C)	grid siderods	36 Ni 64 Fe	Nilvar[1] Carpenter 36[13]
			133	$\pm 0.2 \times 10^{-4}$		73 Ni 20 Cr Al Fe	
55×10^3	20×10^6	25	49	39×10^{-4}	lead wires	28 Ni 18 Co 54 Fe	Therlo[1] Podar[2]
		20	9.0		lead wires, grid lateral wires	50 W 50 Mo	Dowmo[14]
		15	7.6		grid lateral wires	80 Mo 20 W	"H" Wire[14]
		25	8.1		grid lateral wires	75 Mo 25 W	"W" Wire[14]
35×10^3	26×10^6	40	48.2	19×10^{-4}		67 Ni 30 Cu 2 Fe 1 Mn	
45×10^3	26×10^6	40	62.4	19×10^{-4}		66 Ni 29 Cu 2.7 Al	
		30	112	1.5×10^{-4} (20-500 C)		60 Ni 16 Cr Bal. Fe	Tophet C[2]
	31×10^6	30	108	2.2×10^{-4} (20-100 C)		80 Ni 20 Cr	Tophet A[2]
20×10^3	30×10^6	40	9.5	43×10^{-4}	grid siderods, lead wires	99 Ni	
35×10^3	30×10^6	35	18	36×10^{-4}	grid lateral wires	94 Ni 5 Mn	Magno[1]
35×10^3	30×10^6	35	14	45×10^{-4} (20-100 C)	grid siderods	97 Ni 2 Mn	
			27		cathodes, heater wires	58 Ni 39 Co 2 W 1 Al	
			11		cathodes, heater wires	82 Ni 18 Co	
			12		cathodes, heater wires	62 Ni 38 Co	

(Cont'd)

Table XIX (Cont'd)

Properties of Alloys

Alloy•	Density g/cc (at 20 deg C)	Melting Point deg C	Thermal Conductivity watts/cm²/cm/deg C/sec (near 20 C)	Thermal Conductivity cal/cm²/cm/deg C/sec (near 20 C)	Thermal Expansion microinches/deg C (at 20 C)	Specific Heat— cal/g/deg C (at 20 C)	Tensile Strength psi
Permanickel[4]	8.8	1440	0.27	0.06	13	0.13	101×10^3 20×10^3 †
Phosphor Bronze "A" P19*	8.9	1050	0.79	0.19	17.8	0.09	47×10^3 20×10^3 †
Platinum-Rhodium	20	1850	0.30	0.072	13 (0-500 C)		45×10^3
Silver Solder "BT"[7]	10	779					53×10^3
Silver Solder "Easy-Flo"[7]	9.5	627					
Solder 60-40	8.4	183	0.50	0.12 (0-180 C)	24.7 (15-110 C)		29×10^3
Stainless 304 S44*	7.9	1400	0.13	0.032	20 (0-1000 C)	0.12	105×10^3 20×10^3 †
Stainless 430 S78*	7.7	1500	0.17	0.041	9 (0-100 C)	0.11	70×10^3
Steel 1010 S50*	7.9		0.59	0.14	13 (20-100 C)	0.12	47×10^3
52 Alloy[1]	8.3	1425	0.18	0.044	9.5 (20-500 C)	0.115	70×10^3
780 Alloy[15]	7.2		0.08	0.02	20		68×10^3

•Superscript numbers indicate manufacturers of alloys as follows:

1 Driver Harris
2 Wilbur B. Driver
3 Allegheny
4 International Nickel Co.
5 H. K. Porter Co.
6 Westinghouse
7 Handy & Harmon
8 Western Gold & Platinum Co.
9 American Plating Work

*RCA code.

Table XIX (Cont'd)

Properties of Alloys

Yield Strength psi	Modulus of Elasticity psi	Per Cent Maximum Elongation	Electrical Resistivity microhm-cm (at 20 C)	Temperature Coefficient of Resistivity	Present or Potential Use in Electron Tubes	Approximate Composition (per cent)	Other Designations•
50×10^3	30×10^6	35	16.6	36×10^{-4}	grid lateral wires	97 Ni	Nickel "Z" Type B
19×10^3	16×10^6	64	9.6			95 Cu 5 Sn	
		35	19.2	17×10^{-4} (20-100 C)		90 Pt 10 Rh	
		20				72 Ag 28 Cu	Silvaloy 301[9] ML[10] Cusil[8]
						50 Ag 15.5 Cu 16.5 Zn 18 Cd	Silvaloy 50[9] KH-7[10]
			14.5		soldering lead wires to top caps and base pins	63 Sn 37 Pb	
45×10^3	29×10^6	50	70	9.4×10^{-4} (20-500 C)		18 Cr 8 Ni 0.08 C Bal. Fe 2 Mn 1 Si	Nilstain[2] Almet 304[5]
40×10^3	29×10^6	25	60			16 Cr 1 Mn 1 Si Bal. Fe	Almet 430[5] Nilstain 430[2]
26×10^3		28	14.2		lead wires, headers and insert, metal envelopes, plates, grid siderods	0.08-0.13 C 0.3-0.6 Mn Bal. Fe	
20×10^3		33	43.2	36×10^{-4} (20-100 C)	top caps, metal envelopes	51 Ni 49 Fe	Hypernik[6] Conpernik[6] Carpenter 49[13]
24×10^3	14×10^6	35	149			20 Cu 80 Mn	

[10] General Plate
[11] Hoskins Mfg. Co.
[12] Columbia Tool Steel Co.
[13] Carpenter Steel Co.
[14] Radio Corp. of America
[15] Chicago Development Corp.
[16] Welding Equipment & Supply Co.
[17] Haynes Stellite Div., Union Carbide Corp.

†Measured at 800 C.

Table XX

Coefficients of Thermal Expansion at Elevated Temperatures

microinches/inch/degree C

Metal	Temperature—deg C									
	100	200	300	400	500	600	700	800	900	1000
Chromium	6.6	. . .	8.4	. . .	. . .	. . .	9.4	. . .	. . .	. . .
Copper	17.1	17.2	. . .	. . .	18.3	. . .	. . .	. . .	. . .	20.3
Gold	14.2	. . .	. . .	. . .	15.2	. . .	. . .	. . .	16.7	. . .
Iridium	6.8	. . .	. . .	. . .	7.2	. . .	. . .	. . .	. . .	7.8
Iron	12.2	12.9	. . .	13.8	. . .	14.5	. . .	14.6	. . .	. . .
Lead	26.1	27.0	. . .	28.9	. . .	. . .	. . .	. . .	. . .	. . .
Molybdenum	5.2	. . .	. . .	. . .	5.7	. . .	. . .	. . .	. . .	. . .
Nickel	13.3	13.9	14.4	14.8	15.2	. . .	. . .	. . .	16.3	. . .
Palladium	11.1	. . .	. . .	. . .	12.4	. . .	. . .	. . .	. . .	13.6
Platinum	9.1	. . .	. . .	. . .	9.6	. . .	. . .	. . .	. . .	10.2
Rhodium	8.5	. . .	. . .	. . .	9.8	. . .	. . .	. . .	. . .	10.8
Silver	19.6	. . .	. . .	. . .	20.6	. . .	. . .	. . .	22.4	. . .
Tantalum	6.5	. . .	. . .	. . .	6.6	. . .	. . .	. . .	. . .	. . .
Tin	23.8	24.2	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Titanium	8.8	9.1	. . .	9.4	. . .	9.7	. . .	9.9	. . .	. . .
Tungsten	4.5	. . .	4.6	. . .	. . .	5.4 (2000C)	. . .	6.6 (3000C)	. . .	. . .

Table XXI

Tensile Strengths at Elevated Temperatures

thousands of pounds per square inch

Metal	Condition	Temperature--deg C							
		300	400	500	600	700	800	900	1000
Tough pitch copper	Hot rolled—annealed	21.0	16.7	11.5	- - -	- - -	- - -	- - -	- - -
Cartridge brass	Strip annealed	46.0	34.0	13.8	- - -	- - -	- - -	- - -	- - -
Phosphor bronze	Rod annealed	36.8 (250C)	31.0 (350C)	24.4 (450C)	- - -	- - -	- - -	- - -	- - -
Copper/nickel (70/30)	Strip annealed	47.0	42.8	34.4	21.6	- - -	- - -	- - -	- - -
Nickel	Hot rolled	- - -	70	- - -	33	- - -	22	- - -	6
Monel	Hot rolled	- - -	60	- - -	34	- - -	16	- - -	7
K Monel	Hot rolled	- - -	- - -	78	59	- - -	25	- - -	9
Inconel	Hot rolled	- - -	72	- - -	67	- - -	32	- - -	13
Inconel X	Fully heat treated	- - -	- - -	132 (450C)	107.2 (550C)	- - -	46.4	- - -	- - -
Hastelloy B	Hot rolled	- - -	- - -	102	- - -	76	- - -	44	24
Duranickel	Hot rolled—age hardened	- - -	142	- - -	84	17	- - -	- - -	- - -
Stainless 304	Softened	- - -	- - -	- - -	42	- - -	20	- - -	- - -

Table XXII

Total Emissivity

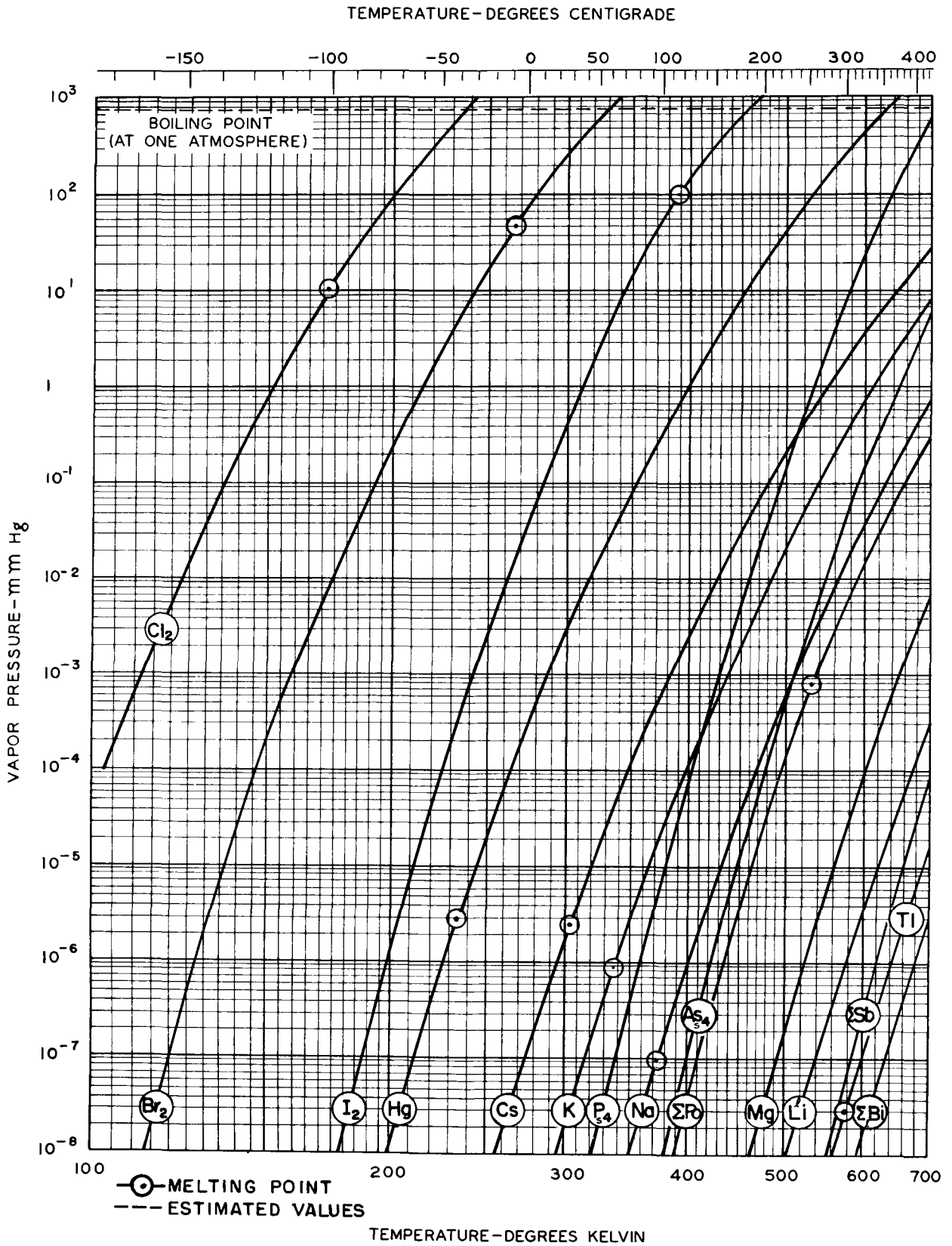
Metal	Emissivity at 100 C	Metal	Emissivity at 100 C
Aluminum (polished)	0.095	Nickel (polished)	0.072
Brass (polished)	0.059	Nickel-Silver (polished)	0.135
Chromium (polished)	0.075	Silver (polished)	0.052
Copper (polished)	0.052	Stainless Steel (polished)	0.11
Copper-Nickel (polished)	0.059	Steel (polished)	0.066
Iron (rough polished)	0.27	Tungsten (polished)	0.066
Molybdenum (polished)	0.071		

Table XXIII

Thermal Conductivity at Elevated Temperatures

Watts/cm^2/cm/deg C/sec

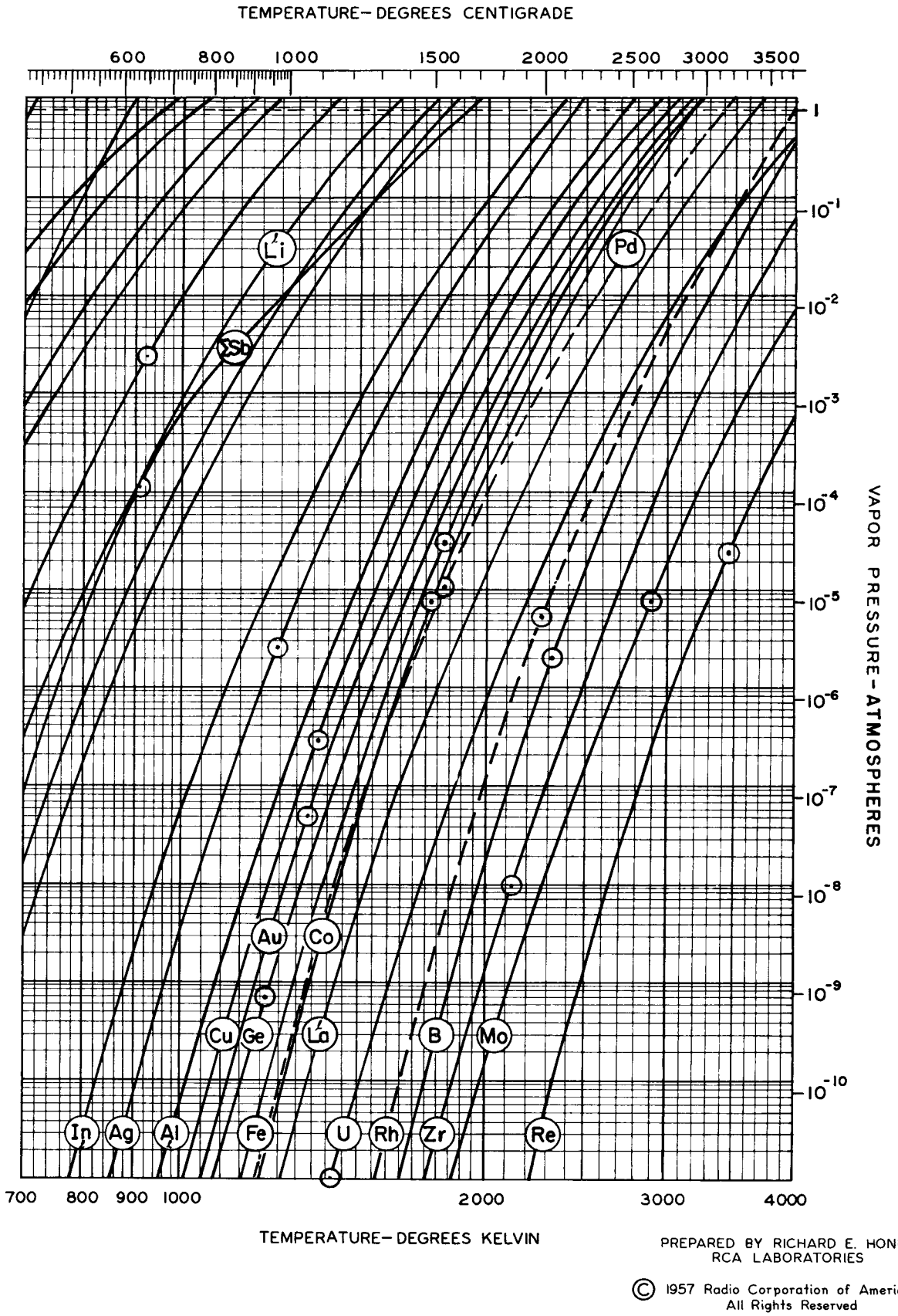
Metal	Temperature--deg C									
	100	200	300	400	500	600	700	800	900	1000
Elementary Metals										
Aluminum	2.18	2.07	1.99	1.92	1.89	- - -	- - -	- - -	- - -	- - -
Copper	3.81	3.77	3.71	3.64	3.60	- - -	- - -	- - -	- - -	- - -
Gold	2.95	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Iron	0.70	0.62	0.56	0.49	0.44	- - -	- - -	- - -	- - -	- - -
Lead	0.33	0.32	0.31	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Magnesium	1.42	1.34	1.30	1.30	1.30	- - -	- - -	- - -	- - -	- - -
Molybdenum	1.38	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Nickel	0.59	0.57	0.54	0.53	0.54	- - -	- - -	- - -	- - -	- - -
Palladium	1.14	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Platinum	0.72	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Rhodium	0.80	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Silver	4.10	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Tantalum	0.96	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Tin	0.61	0.56	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
Titanium	1.52	1.48	1.44	1.40	1.37	1.33	1.29	- - -	- - -	- - -
Tungsten	1.59	- - -	- - -	- - -	- - -	0.96 (1000C)	1.17 (2000C)	- - -	- - -	- - -
Alloys										
Cartridge Brass	1.04	1.10	1.14	1.16	- - -	- - -	- - -	- - -	- - -	
Inconel	0.12	- - -	- - -	0.17	- - -	- - -	- - -	0.24	- - -	
Inconel X	0.16	- - -	- - -	0.23	- - -	- - -	- - -	0.34	- - -	
Karma	0.13	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	
Kovar	0.15	0.15	0.16	0.17	0.18	0.18	0.19	0.20	0.21	
N91	0.44	0.42	0.40	0.38	0.37	0.36	0.38	0.40	0.42	
N97	0.41	0.40	0.38	0.37	0.36	0.34	0.34	0.32	0.30	
Nichrome	0.13	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	
Nichrome V	0.11	0.13	0.15	0.17	0.18	0.20	0.22	0.24	0.26	
Nickel D	0.40	0.35	0.30	0.33	0.37	0.38	0.43	0.46	0.50	
Stainless 304	0.13	0.15	0.17	0.19	0.20	0.22	0.24	0.26	0.27	
Stainless 430	0.22	0.21	0.23	0.23	0.24	0.24	0.24	0.25	0.25	
52 Alloy	0.18	0.19	0.19	0.20	0.20	0.21	0.21	0.22	0.22	



NOTES:

1. For those elements having two or more important gaseous species of known concentrations, total-vapor-pressure curves were obtained by graphically adding the individual curves. Such curves are identified by a Σ preceding the chemical symbol.
2. Where the melting point falls outside the pressure range of the graph, the letter s (solid) or ℓ (liquid) have been included with the chemical symbols.

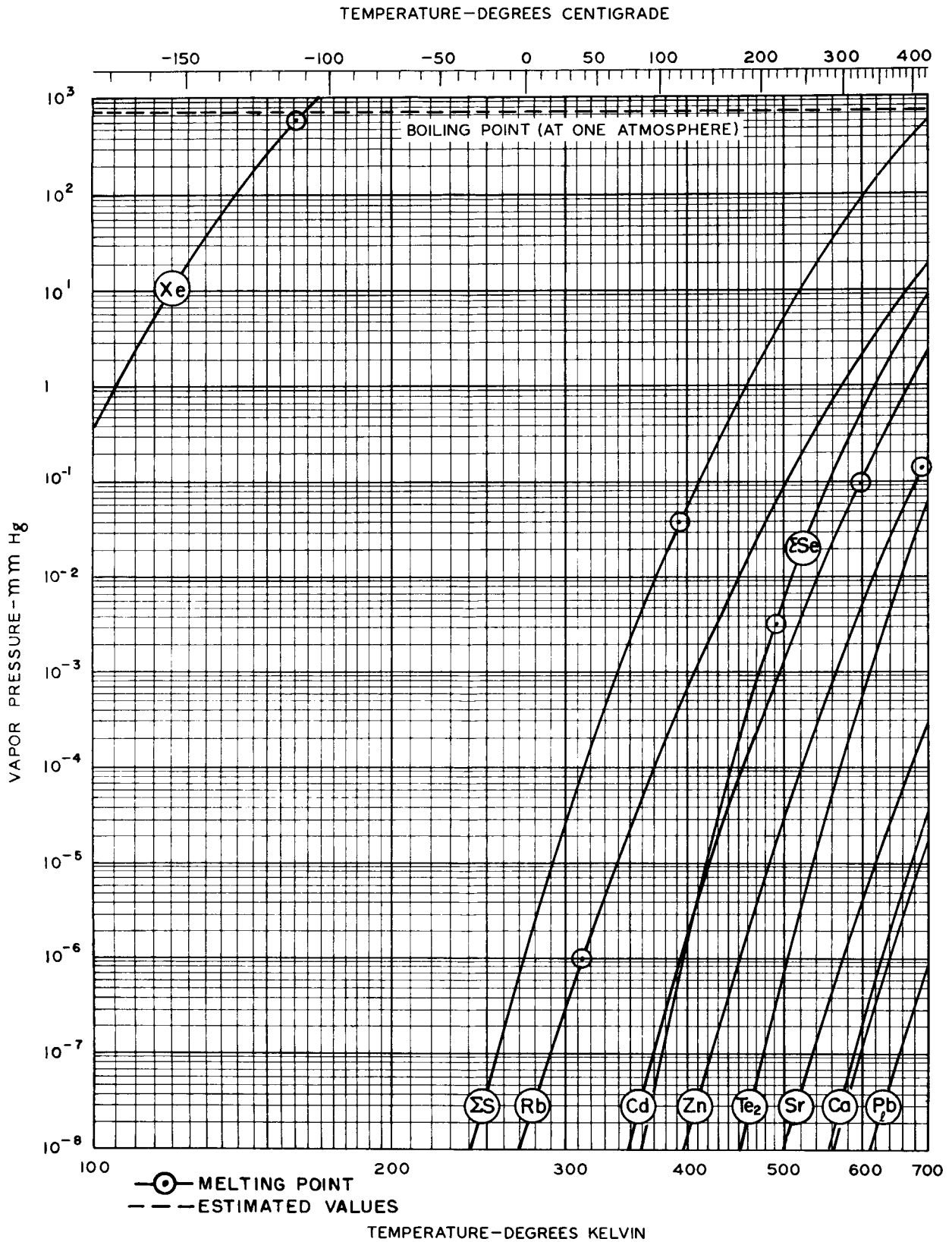
Figure 32. Chart of Vapor Pressures at Elevated Temperatures (I)



© 1957 Radio Corporation of America
All Rights Reserved

3. Where the atomic species is known to be the predominant contributor to the total vapor pressure, contributions from molecular species were neglected, and the chemical symbol has been given without any subscript; for the few elements that consist largely of one molecular species, the appropriate subscript has been added to the chemical symbol.

Figure 32. Chart of Vapor Pressures at Elevated Temperatures (I)

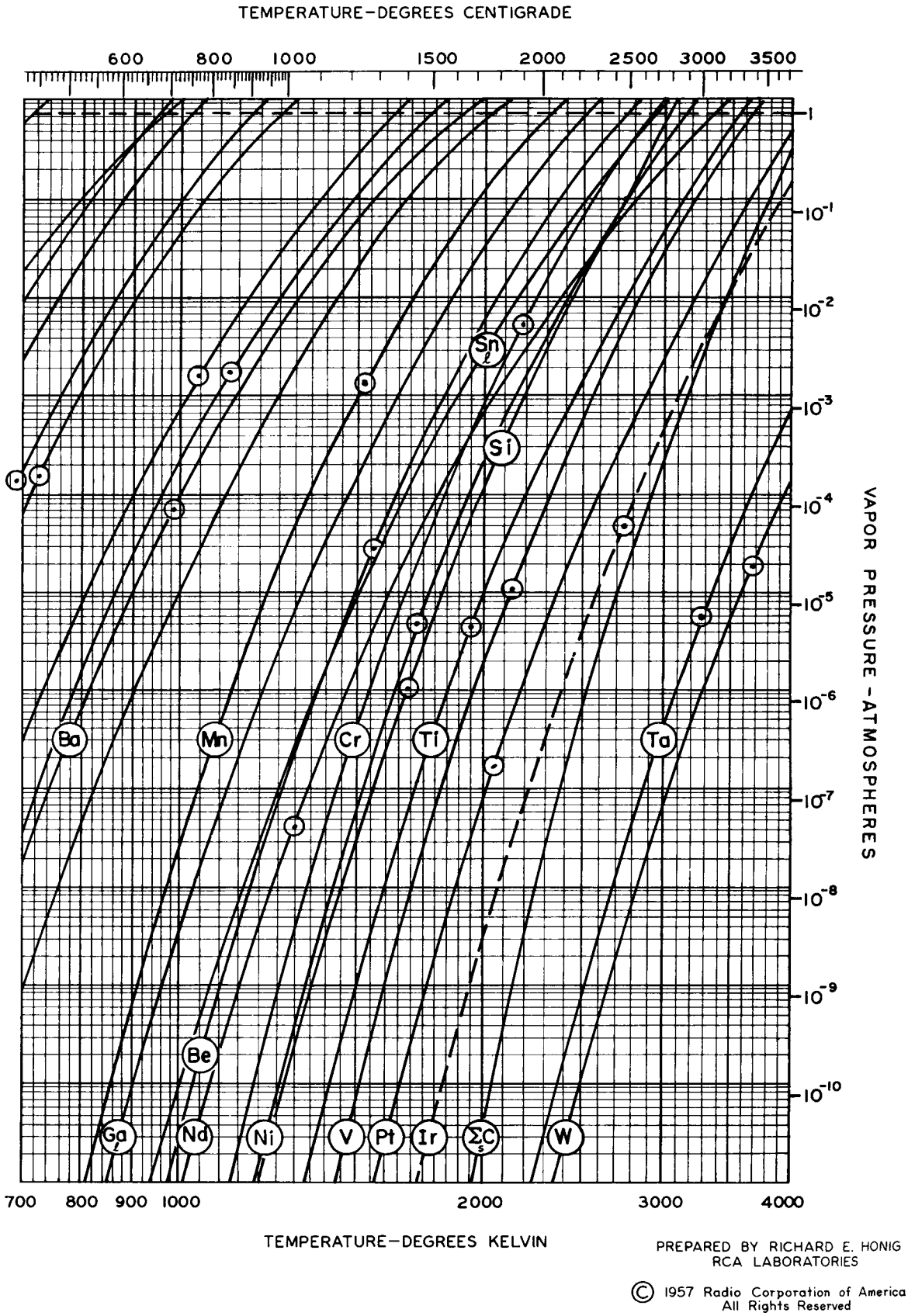


NOTES:

1. For those elements having two or more important gaseous species of known concentrations, total-vapor-pressure curves were obtained by graphically adding the individual curves. Such curves are identified by a Σ preceding the chemical symbol.

2. Where the melting point falls outside the pressure range of the graph, the letter s (solid) or ℓ (liquid) have been included with the chemical symbols.

Figure 33. Chart of Vapor Pressures at Elevated Temperatures (II)



3. Where the atomic species is known to be the predominant contributor to the total vapor pressure, contributions from molecular species were neglected, and the chemical symbol has been given without any subscript; for the few elements that consist largely of one molecular species, the appropriate subscript has been added to the chemical symbol.

Figure 33. Chart of Vapor Pressures at Elevated Temperatures (II)

Metals for Vacuum Tube Construction

P. D. Strubhar

Lancaster

A general discussion of metallurgy and its relation to electron tubes is covered in another article. That article also covers the following subjects: joining metals to metals, joining metals to nonmetals, testing metals, nickel, cathode materials, grid materials, and plate materials. To avoid duplication, the present chapter will be limited to the following subjects: copper, glass-sealing alloys, tungsten, molybdenum, steel, gold, silver, tantalum, titanium, aluminum, and indium. These subjects will be discussed from the viewpoint of tube making in the Lancaster plant.

COPPER

The copper used in the electron-tube industry is not one of the ordinary types of copper which comprise the major part of copper industry production, but is a special grade which is called OHFC* which means "Oxygen-Free High-Conductivity." This copper is made by a special process designed to keep even traces of oxygen out of the copper. The process starts with electrolytic copper, a highly refined copper, which is melted in a furnace containing a protective atmosphere of carbon monoxide made by cracking city gas. This atmosphere is used not only during melting, but also during all phases of casting. This protection is accomplished by the use of hoods over all runners and ingot molds so that the copper is kept in the carbon monoxide atmosphere where no oxygen can come into contact with the copper until it is cool enough to remove it from the ingot molds. These extreme measures to keep oxygen out of the copper are taken because even small traces of oxygen lower the conductivity of the copper very drastically. In addition, small traces of oxygen or copper oxide will combine with hydrogen during the hydrogen-firing processes used to clean tube parts, and may form pockets of steam which make the copper lose its ductility and become very weak.

Figs. 1 and 2 show the effects of hydrogen firing on oxygen-free and oxygen-bearing copper.

American Metals Company, the manufacturer of OFHC copper, makes two grades of this copper—a regular grade and a certified grade. These two grades are made by the same process and it is possible that they could be identical materials. For the regular grade, nominal impurities are given but their amounts are not covered by a warranty. However, if certified material is used, a warranty is given which guarantees

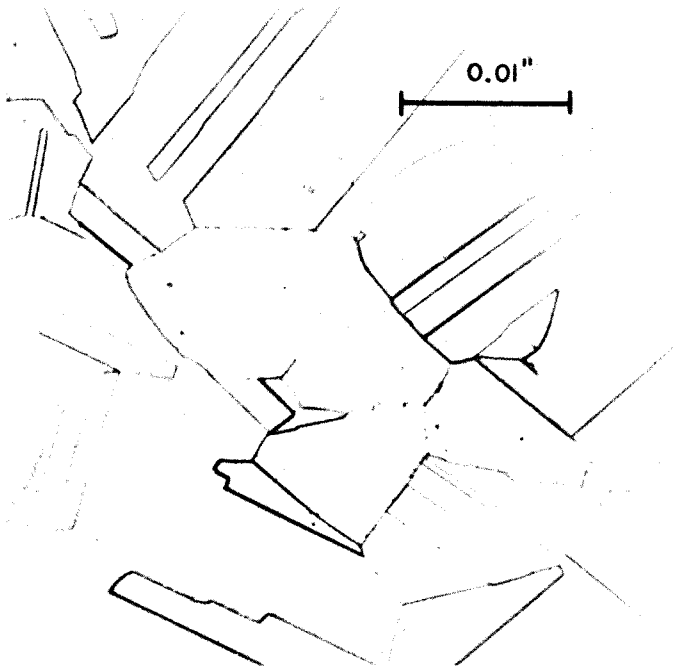


Figure 1. Oxygen-Free Copper After Firing for 10 Minutes in Hydrogen

the impurities to the following limits: lead—0.0010 per cent maximum, phosphorus—0.0003 per cent maximum, mercury—0.0001 per cent maximum, zinc—0.0003 per cent maximum, and sulfur—0.0040 per cent maximum.

Lancaster uses the certified OFHC copper primarily for anodes in power tubes and for exhaust tubulations in both power tubes and cathode-ray tubes. The selection of the certified material is predicated on the fact that the warranty ensures low levels of lead and phosphorus impurities.

The 0.0010 per cent lead maximum is very important when the copper is used in the manufacture of anodes. Many of the tubes using copper anodes are processed at temperatures of over 600 C during exhaust operations. Under these conditions it is essential that the lead content be maintained below the 0.0010 per cent value. Tests have been run which show that if this value for lead goes up to 0.0012 per cent, lead deposits will be formed on the cooler insulating parts of the tube during the exhaust process. Easily vaporized metals such as

* Registered trademark of American Metals Company

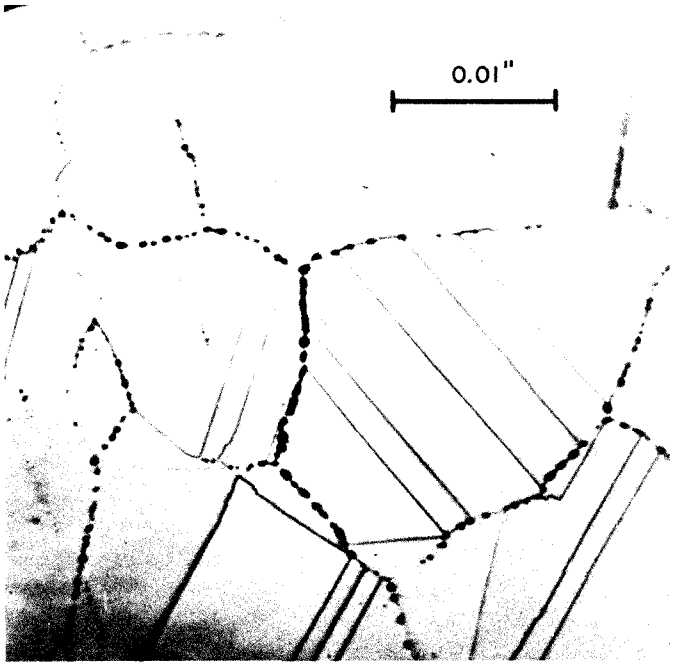


Figure 2. Oxygen-Bearing Copper Showing Voids After Firing at 800 C for 10 Minutes in Hydrogen

mercury and zinc are limited for essentially the same reason. However, because those metals are heated close to or above their boiling points during the manufacture of the copper, a condition not true of lead, the limits for these materials are established more as safeguards against contamination during later processing than as limitations on inherent impurities resulting from the refining operations.

If the copper is used as exhaust tubulations, then the 0.0003 per cent phosphorus limit becomes extremely important. In this use, the exhaust tubulation is "pinched off." Before pinching-off, the exhaust tubulation connects the vacuum tube to the vacuum pumps. After the tube has been completely evacuated, the copper tubulation is compressed between two rollers until it separates into two pieces.

The pinching-off operation removes the tube from the pumps and seals off the tubulation so that the vacuum is maintained in the electron tube. In making the pinch off, the copper is drawn down until it becomes very thin and finally disappears. This leaves a very small section of the copper which is sealed together. (Fig. 3) This operation requires extremely ductile copper. Because the addition of very small amounts of phosphorus has a very marked degrading effect on the ductility of copper, the phosphorus content must be kept very low. If the ductility is not high enough, a small crack will appear in the copper immediately in back of the sealed portion of the pinch off.

An inherent difficulty in the use of copper for tubes is its lack of strength at tube-processing temperatures. Two types of copper alloys are available if more strength is needed. These are a chrome-copper alloy and a nickel-phosphorus-copper (#58 or phosnic bronze) alloy.

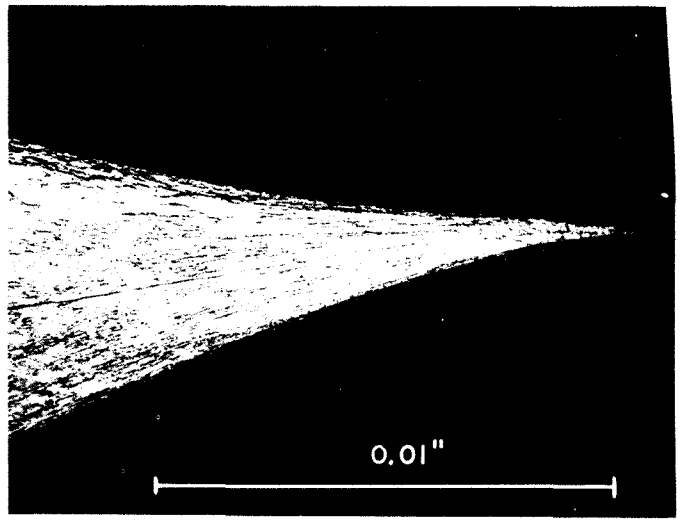


Figure 3. Typical "Pinch Off" Showing Thin Edge at End of Separation

Both of these alloys depend upon their age-hardening property to give them better strength at operating temperatures than is obtainable with the OFHC copper. Practically all other copper alloys made at present cannot be used in tube manufacture because they have detrimental alloying elements, their properties are not generated at tube-processing temperatures, or common joining methods used in tube manufacture cannot be used on them.

GLASS-SEALING METALS AND ALLOYS

In consideration of glass-sealing metals and alloys, two properties become important: (1) the expansion characteristics, and (2) the oxide-forming characteristics. Many types of alloys and glasses are available, but expansion of the metal and of the glass to be joined must be matched to make a successful seal. Whether the matching has to be an absolute match or can be a slight mismatch depends on the type of seal being made. The expansion of the glass should be chosen to keep the glass in compression when the seal is completed. The oxide formed by the metal must be adherent to the metal and reactive to the glass.

Table I gives the expansion limits for the various alloys which Lancaster buys to expansion specifications. Note that all alloys purchased to expansion limits specify only nominal chemical compositions. In specifying expansion limits, it is general practice to specify the temperature range from room temperature to the strain-point temperature of the glass to be used with the alloy. In general, the oxide of the alloy should be adherent to the metal and reactive to the glass. Small deviations from the nominal composition—deviations not large enough to cause the expansion to fall outside the specified limits– do not have much effect on the oxide-forming characteristics of the alloy. However, in many of the alloys, minor impurities can have a very marked effect on the oxide formation. The one element which is limited in practically all glass-sealing alloys is carbon. If the amount of carbon is too high (usually over 0.06 per cent), it causes a seal with an

TABLE I

Coefficient of Expansion Limits

Alloy Composition (nominal per cent)				Coefficient of Expansion	Temperature Range Degrees C
Iron	Nickel	Chromium	Cobalt		
53	29	--	17	5.03— 5.37x10^{-6}	30—450
52	42	6	--	9.60—10.40x10^{-6}	30—400
50	50	--	--	9.50— 9.80x10^{-6}	30—310
72	--	28	--	11.00—11.40x10^{-6}	30—530
83	--	17	--	11.20—11.70x10^{-6}	30—530

unusually large amount of bubbles. At times, materials with a low amount of carbon may cause bubble seals because of carbon left on the surface of the metal from previous operations. Other minor impurities which affect oxide formation can usually be balanced by modifying oxidation procedures. However, in some cases, because of the continual readjustments, it has become necessary to limit some minor impurities. Such limitation has been particularly necessary in the case of the nickel-iron alloys.

After expansion and oxide formation are considered, several further conditions such as extreme temperature transformations and effects of brazing must be considered. At room temperature and above, the alloys which have nickel as an alloying element are in the austenitic crystallographic state. If these materials are subjected to temperatures much below 20 C, they will change to ferrite or an intermediate product. This change will cause an abrupt and localized change in the expansion characteristic and, as a result, the glass will crack. To be sure that these transformation temperatures are not too near to room temperature, samples of these materials are quenched in a mixture of dry-ice and acetone and then examined on a microscope to determine that no transformation has taken place. Alloys with only chromium and iron as the main elements are normally ferritic at room temperature. If these alloys are not properly stabilized, they will change to austenite on glass sealing and thereby cause cracking. As a protection against this happening, these alloys are heated to temperatures slightly above glass-sealing temperatures (1200 C), quenched to room temperature, and examined microscopically for transformation. In making brazes with silver-copper-eutectic solder to the cobalt-nickel-iron alloy, it is necessary to protect this alloy from intergranular attack. The silver-copper-eutectic solder attacks the grain boundaries of this alloy very rapidly when it is stressed (Fig. 4). Therefore, when this alloy is soldered under conditions of stress and thicknesses which are unfavorable, the resultant joint will leak. This reaction can be prevented by plating the alloy with at least a quarter of a mil of nickel (Fig. 5).

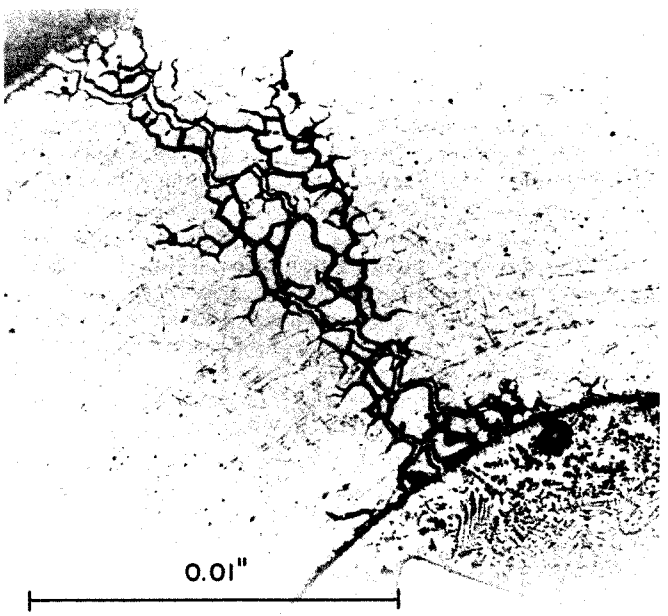


Figure 4. Typical Intergranular Corrosion of Cobalt-Nickel-Iron Alloy by Silver-Copper-Eutectic Alloy

Figure 5. Cobalt-Nickel-Iron Alloy (structure similar to that of Figure 4) Protected by Nickel Plating

In addition to the alloys, there are some metals which are used to make glass-to-metal seals. Tungsten is the most important, with copper, molybdenum, and iron being used to a very limited degree. The expansion characteristics of these metals are given in Table II.

Tungsten, as a glass-sealing metal, is used mainly for

TABLE II

Metal	Coefficient of Expansion Degrees Centigrade	Temperature Range Degrees Centigrade
Tungsten	4.4×10^{-6}	20—300
Copper	17.7×10^{-6}	0—300
Molybdenum	5.5×10^{-6}	20—300
Iron	14.1×10^{-6}	20—500

leads in tubes without a base. It can be sealed without any special atmospheres and, by controlling the mixture of the flames, a strong seal with the proper amount of oxide between the metal and glass can be obtained. It retains high strength even after the sealing operation. The condition of the surface of the tungsten is very important. In the past "splits" had been a serious probelm in using tungsten for leads. Splits are caused by folds from the swaging operation and appear as narrow, deep grooves in the wire. These grooves are so deep and narrow that glass will not flow to the bottom of them and, therefore, the seals will leak. A machine has been developed to detect splits; since it has been in use by the wire manufactures, the problem has not been very serious.

Copper is used to make a seal known as a "Housekeeper" seal which, by making use of the extreme ductility of the copper, can be made with very large mismatch between the glass and the copper. In this seal the copper is thinned down to a feather edge and the glass is allowed to flow down both sides of the copper. The seal is difficult to make and can only be made to OFHC copper or boron-deoxidized copper which have passed a special oxide-adherence test. This test consists of heating the copper to 800 C in air and quenching it in water; if the oxide coating does not flake off, the copper is considered suitable.

Molybdenum is used to make some seals, but because the normal oxide is not adherent to glass, special processes must be used. To make a seal to molybdenum, it is necessary to make the seal in a special atmosphere such as nitrogen or to pretreat the molybdenum with an element such as chromium or silicon. These treatments are designed to form oxides which will adhere to glass. Such adherence is not obtained with molybdenum oxide formed by air firing or by using normal gas-fired sealing flames.

Iron can be used for sealing only if it is pretreated to prevent the formation of a flaky oxide. The most common method of pretreating is to enamel the iron before it is used for the sealing operation. Special precautions must be taken to dissipate the hydrogen absorbed by the iron during enameling.

There has been some investigation into the use of aluminum as a sealing material for glass-to-glass and ceramic-to-ceramic seals. It is possible to use aluminum as a sealing material, but there are a few parameters which must be determined. The advantage of this type of seal is that it can be made with very little distortion of the glass.

Indium is being used successfully for a glass-to-glass seal for the one-half-inch vidicon. This seal can be made at lower temperatures than the aluminum seal, but it cannot be used at as high a temperature as the aluminum seal. The indium seal depends on pressure and some movement of the indium over the glass to make the indium react with the glass; it makes a very good seal without any distortion of the glass.

In addition to the metals and alloys, some composite materials are used for glass sealing. The most outstanding example of a composite material is "dumet." Dumet is used in the form of wire for leads and is made of an alloy core consisting of 42 per cent nickel and 58 per cent iron to which a coating of copper is applied to the extent of approximately 22 per cent of the total cross-sectional area. The copper surface is coated with a thin (0.03-0.30 mils) borate-copper oxide coating which assists in forming the proper amount of oxide in the final seal. The expansion of the dumet is hard to determine. The expansion across the wire matches that of soft glass, while the expansion along the wire is somewhat lower. Normal dumet can be used satisfactorily in seals using wire up to 0.040 inch in diameter. To make consistently good seals with wire over this size, it is necessary to make adjustments in the expansion of the core material. The color of the dumet is the source of much controversy. The color of the borate-copper oxide coating can range from a dark red-purple to a light yellow-red color. There is a partial correlation of the color to the thickness of the oxide. However, although there is good correlation between color and oxide thickness for any one manufacturer, the correlation among manufacturers is poor. Good seals can be made with practically any color of dumet, even without a borate coating, by making the proper adjustments to the sealing fires and making the seal at the proper rate. If there is too much variation in the color of the dumet, good seals cannot be made without adjusting the fires. It is the variation of size of stem, rate of stem formation, and fire adjustments that causes so much controversy. However, anything which helps to give a consistent oxide thickness to the leads as received will help to give a seal with the proper amount of thickness and of oxide solution in the glass of the finished seal.

TUNGSTEN AND MOLYBDENUM

Because the method of making tungsten is similar to that for making molybdenum, their manufacture will be discussed together. Then, the types of material in common use will be discussed. These materials are non-sag tungsten wire, thoriated-tungsten wire, and plain and cross-rolled molybdenum strip.

Both tungsten and molybdenum are made by powder-metallurgy techniques. Because of the high melting temperatures of the metals, the refining and reduction processes are done in the solid state rather than in the liquid state as it is for most other metals. The resulting product is a metallic powder rather than a cast ingot. The powder is then put into a mold and compacted by high pressure into briquettes or ingots, which are very fragile. The briquettes are fired at a high temperature to sinter the particles into one mass. The sintered briquettes have a density of about 80 per cent

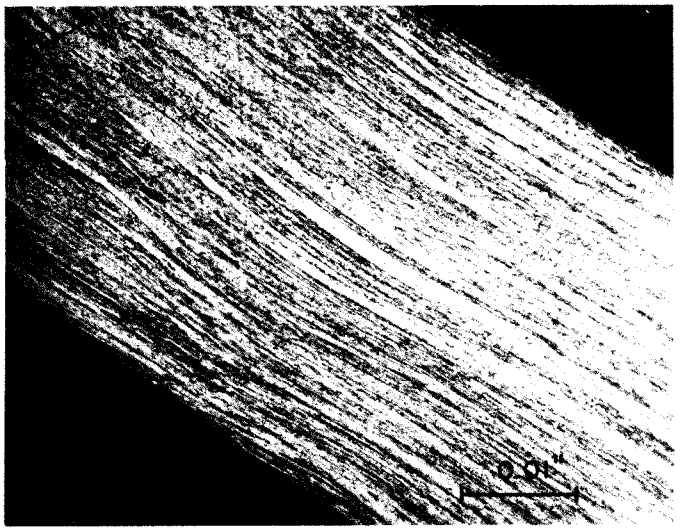


Figure 6. Typical Fibrous Structure of Tungsten Before Recrystallization (The structure of molybdenum is similar)

of the cast material. Sintering, in addition to increasing the density of the briquette, increases its strength markedly. At this stage the only way of forming the metal is by hot working it. However, all heating and working must be down below the recrystallization temperature of the metal or it will again become very brittle. This process results in final material which has an extremely long grained (fibrous) structure (Fig. 6). This material is fairly ductile but will be made very brittle if it is heated above the recrystallization temperature (1200 C to 1700 C for tungsten, 900 C to 1200 C for molybdenum).

Practically all tungsten is used in the form of wire. Two forms of tungsten wire—non-sag and thoriated—will be discussed; both forms are used in the Lancaster plant. Ordinary tungsten wire becomes extremely brittle and badly deformed when recrystallized; recrystallization occurs at the operating temperatures of heaters and filaments. To minimize this effect, it is necessary to "dope" tungsten. The doping is done by mixing alkaline-earth oxides with the powdered tungsten before it is compacted into a briquette. Doping has the effect of inhibiting grain growth during recrystallization; because it is distributed preferentially along the axis of the wire, the doping material, or dope, has more inhibiting effect across the wire than along the wire. This difference causes the grains to grow faster along the wire than across it. Tungsten wire, with this type of structure, while still fairly brittle, is not nearly as brittle as wire without it. Because the dope can be removed by improper processing, there is always the possibility of getting wire which is supposed to be non-sag wire, but which has lost its non-sag properties. Several tests for "non-sag" have been devised, such as (1) making a cantilever loop of the wire and measuring the amount of lowering of the unsupported and when the wire is heated to 2600 K, (2) heating the wire to 2600 K and then observing microscopically the ratio of the length to width of the recrystallized grain, and (3) observing whether recrystallization of the wire takes place at a selected fraction of the maximum current the wire will carry. In the use of tungsten wire, the rate of heating the wire after it passes the recrystallization temperature for the first time is very important. Wire which is heated rapidly (within 20 seconds) from the recrystallization temperature to the operating temperature for the first time will tend to have ordinary tungsten wire properties instead of non-sag properties. To obtain maximum non-sag properties, it should take at least 90 seconds to raise the wire temperature from the recrystallization temperature to the operating temperature. Non-sag wire is usually used for heaters or filaments.

Figs. 7 and 8 show samples of tungsten that have been heated at different rates.

If thoria is added to tungsten wire it makes the wire suitable for use as a cathode in some tubes. The addition of thoria to the tungsten oxide powder has an effect on grain growth somewhat similar to that of doping, but because the thoria is radioactive, its presence can be detected by Geiger-Muller counters; it is not necessary, therefore, to make the various sag tests to determine if thoria is present. Adding thoria to tungsten has the effect of markedly increasing the electron emission allowing the tungsten wire to be used as a cathode. If thoriated-tungsten wire is used as received, the thoria will dissipate itself very rapidly and the emission life will be short. To slow down the thoria dissipation, thoriated-tungsten filaments are carburized. The process consists of heating the filament in a hydrogen atmosphere that has been enriched with a hydrocarbon by bubbling the hydrogen through a liquid hydrocarbon, such as benzene, and of producing a layer of tungsten carbide on the surface of the wire (Fig. 9). It has been shown that the inhibiting action is caused by the fact that the thoria will only migrate along the carbide grain boundaries. Therefore, the type and grain size of the carbide structure become very important factors in determining the life of tubes using carburized thoriated-tungsten filaments. However, there usually must be a compromise between initial properties and extended-life properties and, therefore, the most suitable struc-

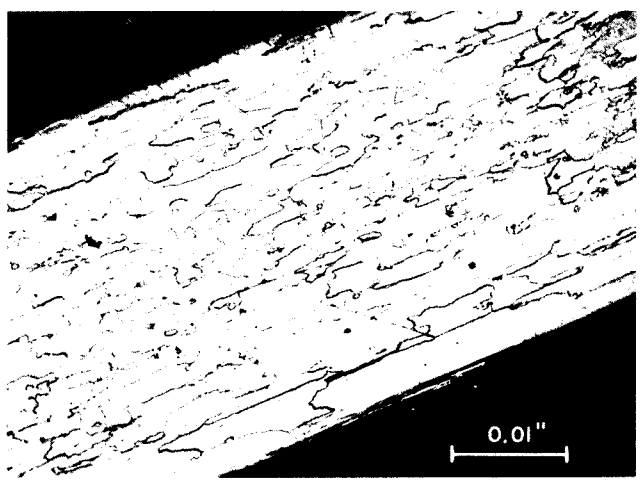


Figure 7. Tungsten Wire Heated From Recrystallization Temperature to 2600 K in 20 Seconds

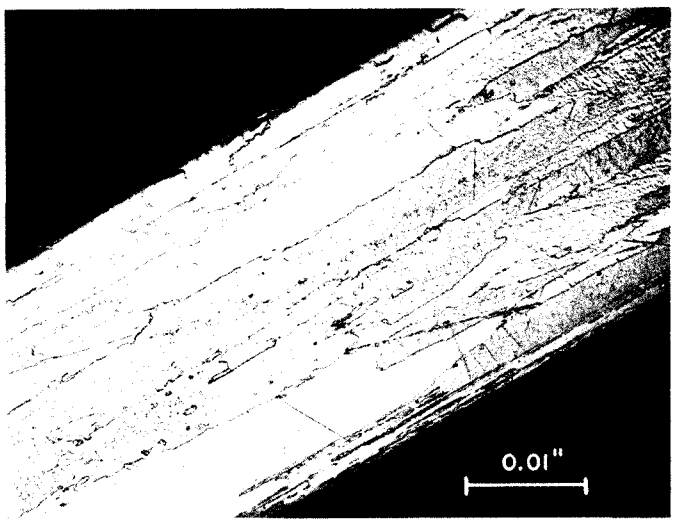


Figure 8. Same Tungsten Wire as Figure 7, but Heated From Recrystallization Temperature to 2600 K in 90 Seconds

ture must be worked out experimentally for each tube type. There are many times when thoriated-tungsten wire does not develop the desirable emission properties under normal processing schedules. Until recently the reasons for such variations has been a complete mystery. Some recent work has shown that the presence of over 0.03 per cent calcium has a very detrimental effect on the life-emission characteristics of tubes with throiated-tungsten filaments. While this information represents a great step forward in understanding some of the idosyncrasies of this material, there are still some other variations which must be investigated.

Molybdenum wire is usually not used at the high temperatures at which tungsten is used and, therefore, recrystallization problems do not appear. However, because fairly large amounts of molybdenum strip are used, some of the problems connected with its use will be discussed. When strip is made from most metals which recrystallize during processing, a product with the same physical properties in all directions is usually obtained. This result is not true with molybdenum. If the material is continually elongated in one direction and not the other, the final product will have very large differences in ductility depending upon the direction of testing. As a result, bends of 180 degrees can be made across the direction of rolling, but very often the material will break when it is bent only 30 degrees along the direction of rolling. To overcome this difficulty, it is necessary to elongate the grains equally in both directions. Material which has been rolled in both directions to equalize grain elongation is called "cross-rolled molybdenum." This type of material should be used if parts are to be made which require bending in several directions or drawing. Cross-rolled molybdenum, because of the special processing and handling, costs much more than regular molybdenum and should be used only when it is absolutely necessary. Raising the temperature of molybdenum to 200 C during forming operations has a marked beneficial effect on the amount of work that can be done on molybdenum. Many parts that cannot be formed at room temperature can be made if the molybdenum and dies are heated during the forming operation. Some recent developments in making a cast molybdenum indicate that many of the problems attendant to the sintered molybdenum can be eliminated by casting techniques; when the casting techniques are ready for commercial use, they should make molybdenum usable for many more parts.

STEEL

In the types of tubes made in the Lancaster plant, a plain low-carbon steel and several types of stainless steels are used. The chromium-type stainless steels have been discussed under glass-sealing alloys; only the nickel-chromium types (302, 304, and 305) will be discussed here.

The plain low-carbon-type steel is a specially selected basic-open-hearth low-carbon steel. It is used to make the shadow mask for the color kinescope. Since the use of this material is low compared to the production rates of steel and because the quality requirements for it are much higher than those required for conventional commercial practice, this material cannot be bought from regular steel mills, but must be selected from the incoming stock of a rerolling or specialty company which then rolls this material to extremely tight mechanical tolerances. This selection is done by taking samples of both ends of a coil of strip. The surface of the samples is checked for defects by etching the surface. Longitudinal sections of the samples are examined for inclusion count, grain size, and carbon distribution. If the sample passes all of these tests, the coil is released for processing into material of the proper size for making masks by the photoengraving techniques.

The most-used type of stainless steel is the 18 per cent chromium--12 per cent nickel (type 305) stainless steel. The main reason for preferring this type of stainless steel to the more common 18 per cent chromium—8 per cent nickel (types 302 and 304) is that it is easier to maintain a lower magnetic permeability with the type 305 material. Since this material is used in gun structures for cathode-ray tubes, magnetic perme-

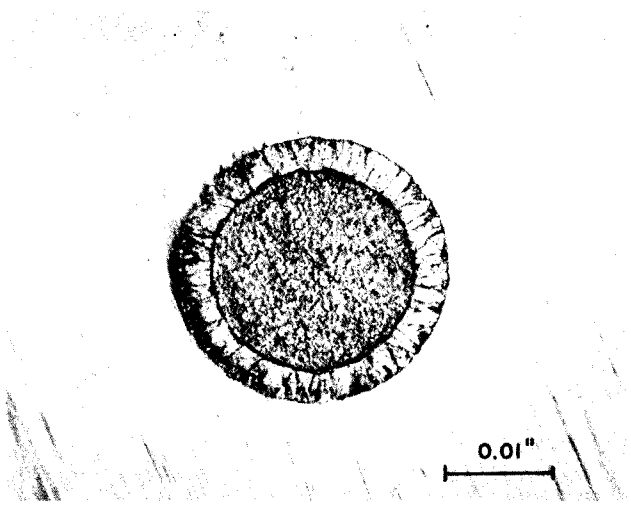


Figure 9. Cross-Section of Tungsten Wire After Carburization (Outer ring is the carbide structure)

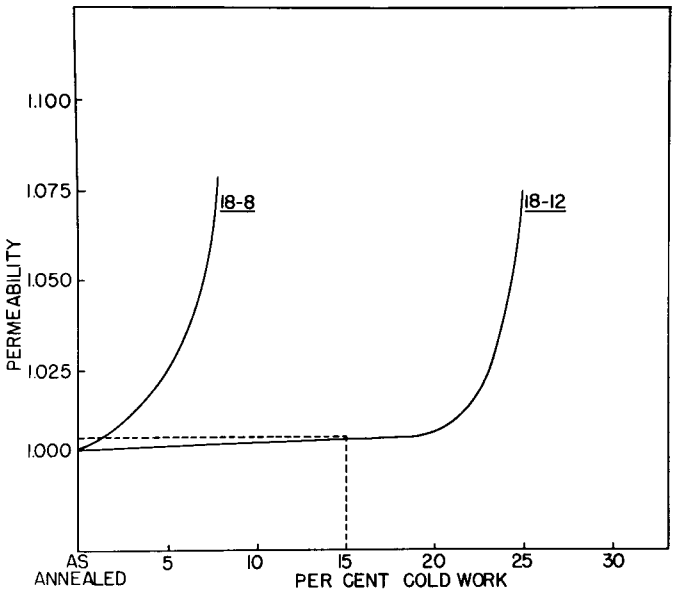


Figure 10. Relationship of Permeability to Cold Work for Two Types of Stainless Steel

ability is an important factor. After both types of stainless steel are annealed, their magnetic permeability will be below the limit of 1.005, but the 12 per cent nickel material will not lose its low permeability nearly as rapidly as the other type when it is cold worked or welded (Fig. 10). This characteristic is important for operations that must be performed after the last anneal. In addition, the 12 per cent nickel type of stainless forms its oxide at slower rates; therefore, it gives a cleaner material for tube construction. Many of the parts used in guns for cathode-ray tubes have deep-drawing operations in their manufacture. Until recently, these parts were at times difficult to make because the vendors would not guarantee a minimum grain size to the material. However, vendors have recently started to sell stainless to a grain-size specification; buying to this specification should reduce the problems connected with the deep-drawing operations of stainless-steel parts. If welding operations are to be performed on stainless-steel parts, the carbon content of the steel should be maintained below 0.08 per cent. If the carbon content goes above this amount, the material around the weld will lose its corrosion resisting properties unless it is re-annealed after welding.

OTHER METALS

Metals such as gold, silver, tantalum, titanium, aluminum, and some of their alloys are not used very extensively but they should be referred to and some pertinent facts pointed out about them. These materials may be rather expensive and therefore are used only when they provide some special property which is needed and which cannot be obtained with the more common metals.

Most of the gold used in the Lancaster plant is either plated or clad on other materials. Practically no gold is used in wrought form. The primary use of gold is to prevent secondary emission. Because gold is very weak structurally in its pure form, it is usually put over other materials which provide the necessary structural strength.

On the other hand, a silver alloy—2 per cent magnesium silver—is used in multiplier phototubes and image orthicons for its high secondary emission. This alloy requires extreme care in handling because the magnesium must be in the metallic form rather than the oxide form. Because magnesium forms its oxide very readily, this requirement presents some problems in melting and working the alloy. After the melting and hot-working of the alloy, large amounts of it must be removed from the surface of the slab to get below the oxidized magnesium to the pure metal. Subsequently, all heat treating must be done in a vacuum of less than 10^{-5} millimeters of mercury to prevent oxidation of the magnesium. The material is oxidized by a controlled process in tube processing to give it excellent secondary-emission properties. Some pure silver is used as dynodes in image orthicons. The material is of very high purity (99.99 per cent pure) and is used because of its special emission properties which can be developed by tube-processing techniques.

Tantalum, in strip form, is used where an extremely high-melting-point material which can be easily formed into complicated shapes is needed. Tantalum is very ductile at room temperature and can be made into parts which must withstand high temperatures. At elevated temperatures, it is not as stiff or as strong as tungsten or molybdenum, but where complicated parts are needed that cannot be formed from these metals, tantalum is an excellent substitute.

Titanium is an element which has been receiving promotion recently. It has some gettering ability, but is not as desirable as the presently used getters unless gettering action is needed for hydrogen or nitrogen. It has two serious disadvantages—its hot strength is very low and its cost is very high. For these reasons, very little work has been done with titanium.

Aluminum, in wrought forms, has very little application in the Lancaster plant. Many attempts have been made to use the aluminum-clad steels in small power tubes. These have always resulted in failure. The processing temperatures are so high that invariably the aluminum has deposited on the cathode and caused poor emission. A vaporized film of aluminum is used to back up the phosphors in cathode-ray tubes to increase the light output of these tubes by reflecting the light from the phosphors toward the viewer.

Principles of Metal Forming

L. S. Sloane

Lancaster

The forming of metals covers a very large field. Rather than attempt to describe all of the forming operations such as forging, swaging, and wire drawing, this chapter will be limited to mechanical press and four-slide machine operations only. These machines are the most widely used in the manufacture of electron tube components. Most of the parts used in industrial and entertainment tubes are stampings; some of the larger power tubes use drawn shells as well as stampings.

The more specialized operations such as rolling and hydroforming are so limited in application that discussion would be of little benefit to the tube engineer. Final operations, such as those performed on lathes, drill presses, and milling machines, although necessary in the manufacture of tube components, will not be discussed because it is considered that these operations are basic and that the engineer has undoubtedly had some experience with them.

This chapter is intended only to familiarize the engineer with the various aspects of metal-forming operations. No attempt is made to make the tube engineer a tool designer. The tube engineer should always consult with the tool designer before component designs are finally decided upon. Only then will it be possible to produce the most economical part on the least expensive tool.

THE MECHANICAL PRESS

Mechanical presses are usually classified by their source of power, frame type, bed type, and means of actuation of slides. There are two major types of mechanical presses: the flywheel type, or plain press, and the geared type. Other characteristics of a press are the capacity (the force exerted near the bottom of its stroke), the stroke of the machine (the distance the ram travels in a vertical direction), the slide adjustment (the distance the slide can be moved in relation to the bed), and the shut height (the distance from the top of the bed to the bottom of the slide with the stroke down and adjustment up).

The flywheel type of press transmits the energy of the motor directly to the flywheel by means of belts. The flywheel is mounted on the main shaft or eccentric. This type of press is normally used for light stampings or other light high-speed operations.

The geared type transmits the energy of the motor to the flywheel through a gear reduction, or a series of gear reductions, to the main shaft. Presses of this type are used for heavier stampings and drawing operations. They generally run slower than the flywheel types but their power is more positive with less slowdown of the flywheel during the work period.

Typical presses are of many forms but the most widely used press is the open-back inclinable (Fig. 1).

The action of a press can either be single, double, or triple. The single-action press is one which has only one slide. The double-action press has two slides, the outer slide carries the blankholder and the inner slide carries the punch. The outer slide usually has a shorter stroke and dwells while the inner slide continues to descend to perform the drawing operation. Three slides are utilized in the triple-action presses having three motions properly synchronized for drawing, redrawing, and forming. The blankholder and plunger slides are located above the bed and the lower slide is located within the bed. The upper slides act in the same manner as in the double-action presses except that the inner slide dwells briefly at the bottom of its stroke while the lower slide performs its function. The lower slide is usually actuated by an eccentric or crank.

The single-action press can be used for operations generally performed in double- and triple-action presses by use of an air cushion mounted under the bed of the press and dies using springs and other mechanical means to apply the necessary pressures.

PRESS OPERATIONS

Numerous operations are performed by mechanical presses. Therefore, for the sake of simplicity, only the basic operations — blanking, piercing, drawing, coining, extruding, and forming will be discussed.

BLANKING AND PIERCING

A blanking or piercing operation is one in which material is sheared from stock. Blanking consists of stamping the outer form of a part; piercing is the stamping of holes within the part. In either operation, the metal is stressed in shear between two cutting edges to beyond its ultimate strength. As the pressure of the punch is applied, tensile and compressive stresses

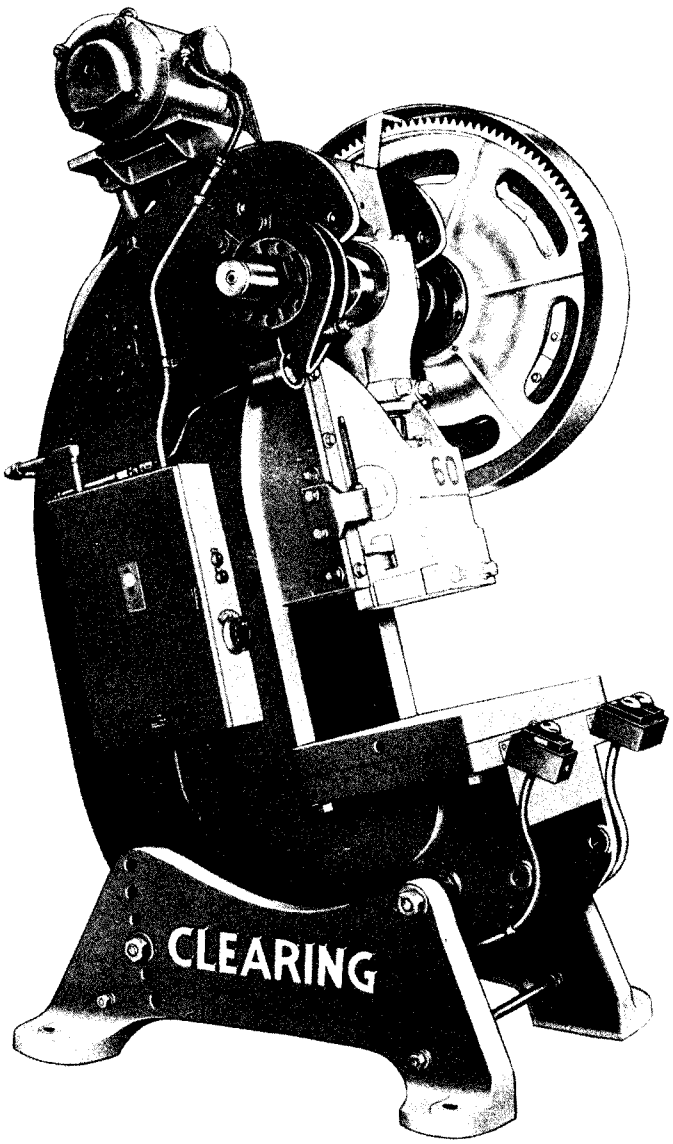


Figure 1. Open Back Inclinable Press

are induced in the material. The pressure increases until the material finally exceeds its elastic limit and it is forced into the die opening. As further pressure is applied, the punch penetrates the material causing a reduction in the cross-sectional area of the metal. Fractures, which will start in the reduced area at both the upper and lower cutting edges, will spread toward each other as pressure continues to be applied until the fractures meet and complete separation occurs. Fig. 2 graphically demonstrates this action.

Clearance between the punch and die is extremely important to insure a clean appearance of the blank edge or pierced hole. The clearance required is a function of the characteristics of the material to be stamped and varies from about 4 per cent to 8 per cent of the thickness of the material. When the correct clearance is used, the punch penetrates the metal to a depth of about one third the metal thickness and forces an equal portion of the metal into the die opening before fracture occurs. The portion penetrated will appear as a highly burnished band around the periphery of the blank or hole.

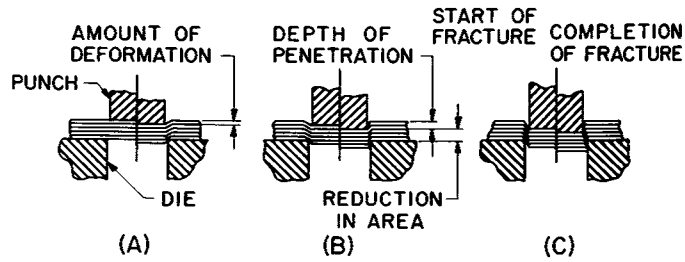


Figure 2. The Shearing of Metal: (A) plastic deformation, (B) reduction in area, (C) fracture

Excessive clearance will cause a larger radius on the edge of the material where the punch is applied and a burr on the opposite edge; insufficient clearance will produce additional burnished bands and the material between and below the bands will be rough on both the pierced hole and the blank or slug (Fig. 3).

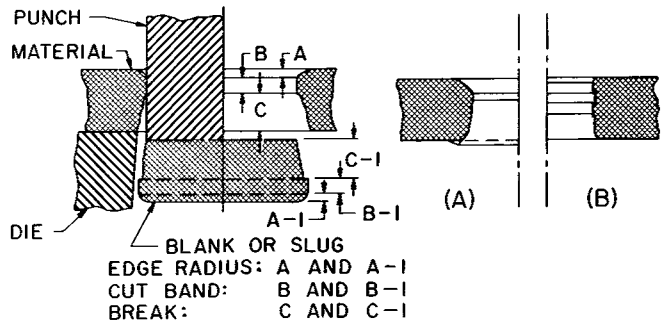


Figure 3. Effect of Die Clearance: (A) excessive clearance, (B) insufficient clearance

The width of the burnished band is usually an indication of the hardness of the material; the wider the band, the softer the material. Harder materials require larger clearances, but permit less penetration by the punch than ductile materials. Dull tools give the effect of insufficient clearance and produce burrs on the edges of the part.

In a blanking die, the die block is ground to the size of the part desired and the punch is reduced by twice the clearance. Piercing dies are made so that the punch is ground to the size of the hole required and the die block is increased by twice the clearance. Fig. 4 illustrates a typical blanking or piercing die.

DRAWING

Drawing is a process of cold forming whereby a flat blank is forced into a die cavity by means of a punch. In operation, the blank is placed over a pressure ring which is usually, but not necessarily, mounted on the lower die set. Then, as the slide starts its descent, the upper die contacts the blank and restrains it while the material is forced over the punch as shown in Fig. 5.

Blankholding Pressure

Confining the material between the upper die and the pressure ring causes the material to flow over the punch

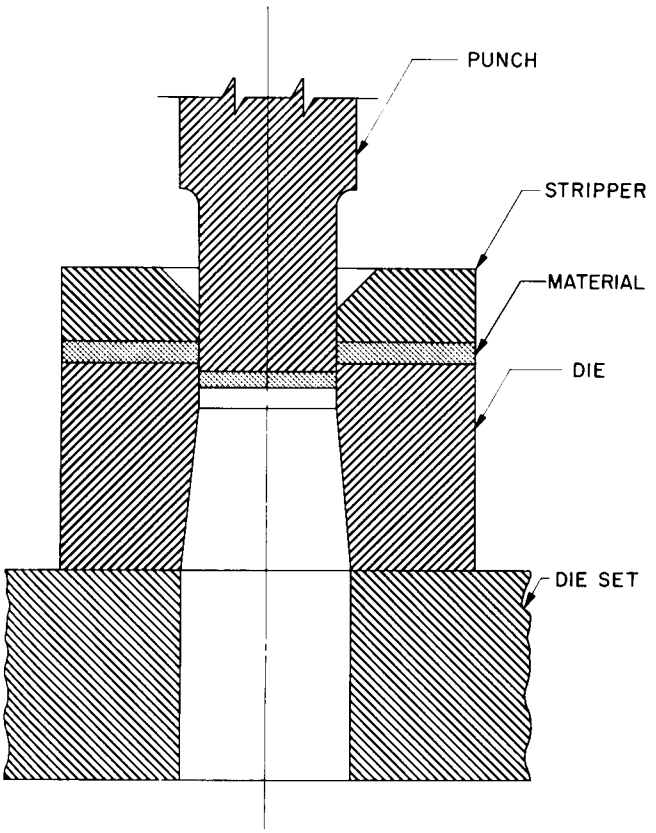


Figure 4. Blanking or Piercing Die

and prevents wrinkling of the material. This action can be easily demonstrated by placing a rubber sheet over a block with a hole in it. If the sheet is merely forced through the hole, the sheet will wrinkle and bunch up as it is pushed into the cavity. However, by placing another block with a hole in it (pressure ring) over the sheeting and applying slight pressure to the block, the sheet will flow into the cavity without wrinkling.

The pressure can be applied to the draw dies in several ways, pneumatic or hydraulic cushions, springs, or rubber blocks, can be used. Cushions are mounted below the bed of the press while springs and rubber blocks are built into the die set. The pneumatic-type cushion is the most widely used. It is merely an air cylinder whose pressure increases as the top of the cylinder is depressed. Pressure pins connect the pressure ring to the cushion. Starting with some predetermined pressure in the cushion, the pressure is increased as the ram descends and the material is forced over the punch. Insufficient pressure will cause wrinkling; excessive pressure will prevent plastic flow and will result in the punch being forced through the material.

Draw Radius

The radius on the die block is extremely important in obtaining the proper flow of material during the drawing process. Insufficient radius will restrain the material flow and the punch will have a tendency to push through the material — an effect similar to that of excessive blankholding pressure. This radius also directs the flow: the material will conform to the punch when

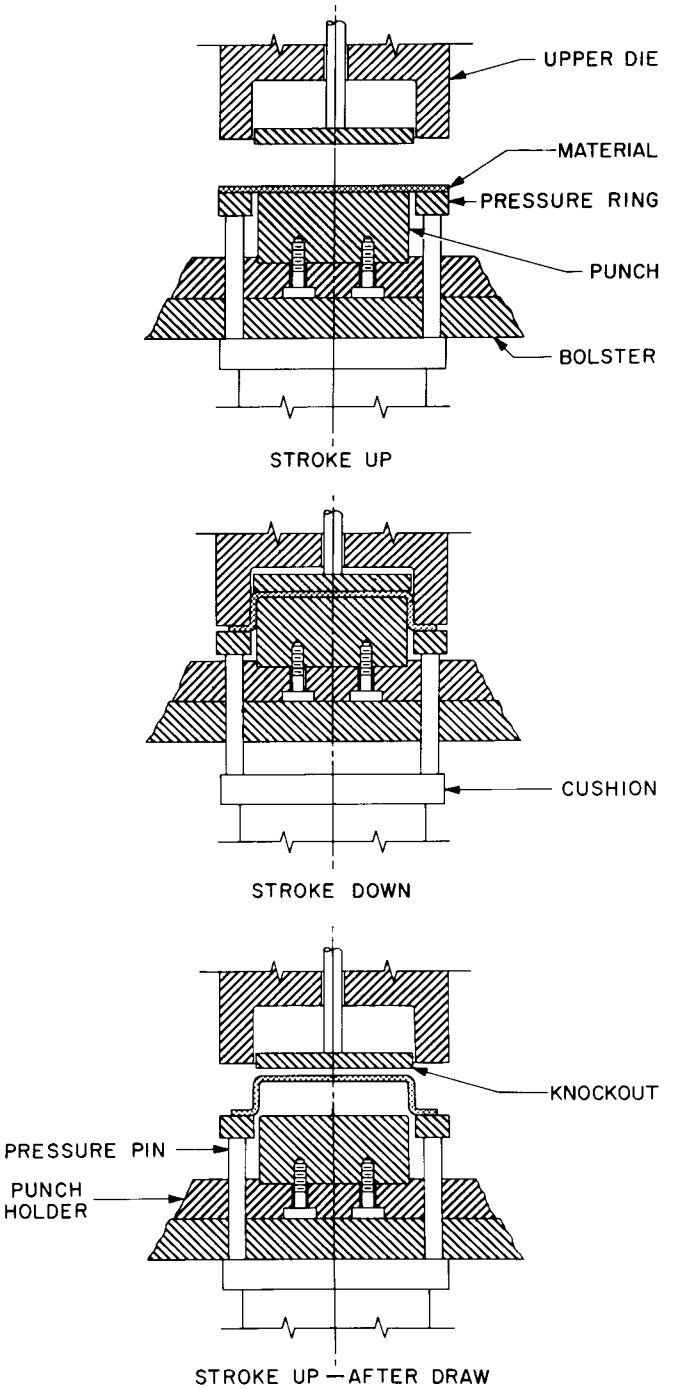


Figure 5. Operation of Draw Die

the radius is small; a large radius will force the material to take the size of the die block. Therefore, if the inside diameter of the drawn cup is to be maintained to close tolerances, the die block radius is made smaller. Conversely, if the outside diameter of the part must be held, the die block radius is made larger.

In practice, the radii are usually at least twice the material thickness; design engineers should endeavor to standardize all inside radii by this rule of thumb. Smaller radii, however, can be obtained by a sizing operation whereby the part is drawn with an oversized radius and then redrawn to the specified dimension.

Material Reduction

Material drawn over a punch will stretch and thereby reduce the thickness of the material. The greatest reduction occurs above the radius at the closed end of the cup. The wall thickness increases from that point to the open end of the cup. The wall thickness at the open end is always greater than the original thickness; the material at the bottom of the cup has nearly the same thickness as the original stock. Fig. 6 illustrates this thickening and thinning of drawn material which had an original thickness of 0.085 inch.

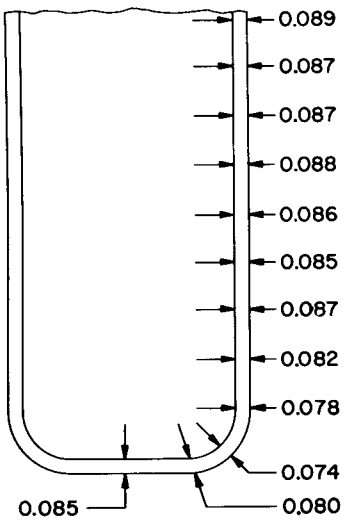


Figure 6. Wall Thickness Variation After Drawing (original thickness: 0.085 inch)

It is extremely difficult to determine accurately the amount of material reduction in drawing because the reduction depends on a number of factors: the percentage of draw reduction (which will be explained later), draw radii, pressure required to draw, the number of anneals required to produce the part, and the characteristics of the material. The engineer should make every effort to avoid specifying both inside and outside diameters or the wall thickness unless the tolerances are large enough to allow for thickness changes.

Diameter Reduction

Due to material characteristics, there is a limit to the amount of diameter reduction than can be obtained in one operation. The tool engineer must calculate the blank size required to make the part. From the relation between blank diameter and finished cup diameter, the number of draws necessary to produce the part satisfactorily can be determined. Each draw requires a specific ratio of the area of the material held by the blankholder to the area of the punch. The greatest reduction is always made on the first draw, usually from 35 per cent to 45 per cent. Each succeeding draw, or redraw, is generally kept to about a 25 per cent reduction. A typical diameter reduction of a drawn cup is shown in Fig. 7

Blank Diameter

The approximate blank diameter of symmetrical shells can be determined by calculation, by graphical layout, or by a combination of the two. The factors affecting the thickening and thinning of the wall thickness also affect the determination of the blank diameter. As a result, determinations are only approximate.

In general, the area of a drawn shell will be equal to the area of the blank from which it is drawn. The diameter of the blank required can, therefore, be determined by calculating the area of the shell and determining the diameter of the circular blank that will have the same area.

For flat-bottomed cylindrical shells, the diameter of the blank required is given by the following equation:

$$D = \sqrt{d^2 + 4dh}$$

where D = blank diameter
d = shell diameter
h = shell height

For more intricate shells, the area of the shell is determined by dividing it into simple elements such as annuli, discs, right circular cylinders, and tori. The diameter of the blank will be:

$$D = \sqrt{\frac{4A}{\pi}}$$

where A = the total area of the shell

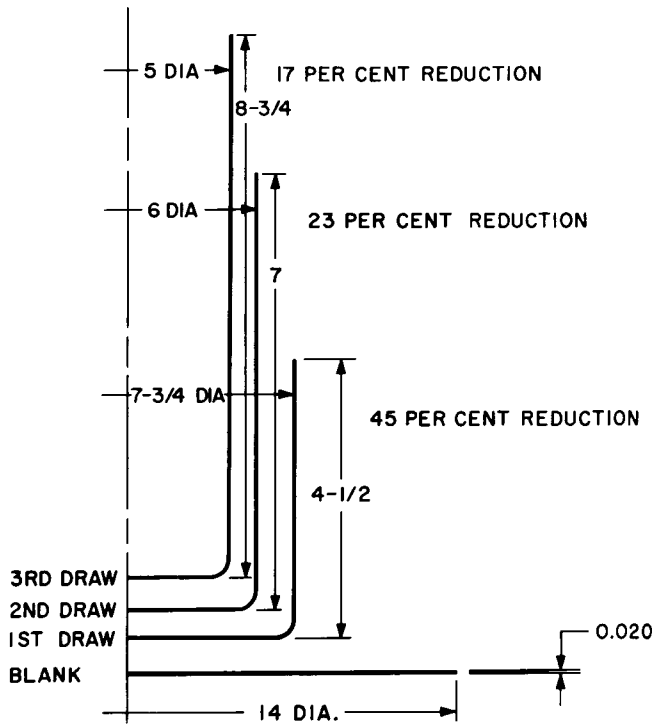


Figure 7. Analysis of Redrawing Operations

Reverse Drawing

Reverse drawing is a redraw operation performed in a direction opposite to the original draw, or drawing the part "inside out." This type of drawing does not

work the metal as severely as ordinary draws and will not wrinkle the part. Fig. 8 shows a typical example of a reverse-draw operation.

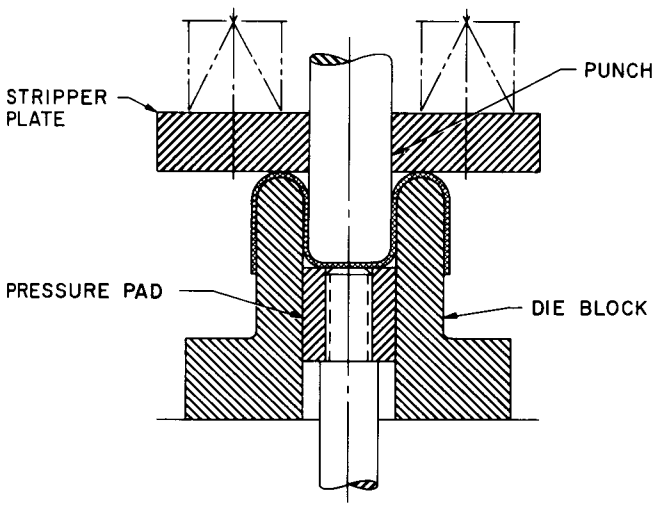


Figure 8. Reverse-Draw Die

SIZING

Sizing is an operation used to obtain accurate dimensions. Close diametral tolerances may be obtained by ironing the sides of the part; the radius at the bottom of the cup may be made smaller by sizing.

COINING

Coining is a squeezing operation in which all the surfaces of the material are confined. In the coining of an aperture, for example, the blank with a hole slightly larger than the desired diameter is placed over a pin. As the coining punch strikes the part, the material is forced to flow around the pin. This same squeezing action is used to obtain square corners on drawn cups.

EXTRUDING

Extruding is a squeezing action, but unlike coining, the extruding die directs the flow of material ahead of or behind the extruding punch. When the material is directed ahead of the punch, the operation is known as forward extrusion; when it is directed behind the punch, the operation is known as backward extrusion. (See Fig. 9.)

FORMING

A wide variety of forming tools are used in the manufacture of tube components. Wire supports, leads, filaments, plates, shields, and radiators, are only a few of the parts made by forming operations.

The biggest problem involved in the forming of metals is the "springback" of the material. Simple forming tools consist of a die block and a forming punch. The tool is usually ground to counteract the effect of springback, but in many cases springback cannot be controlled because it varies greatly within each individual coil of material and between coils of the same material. As a

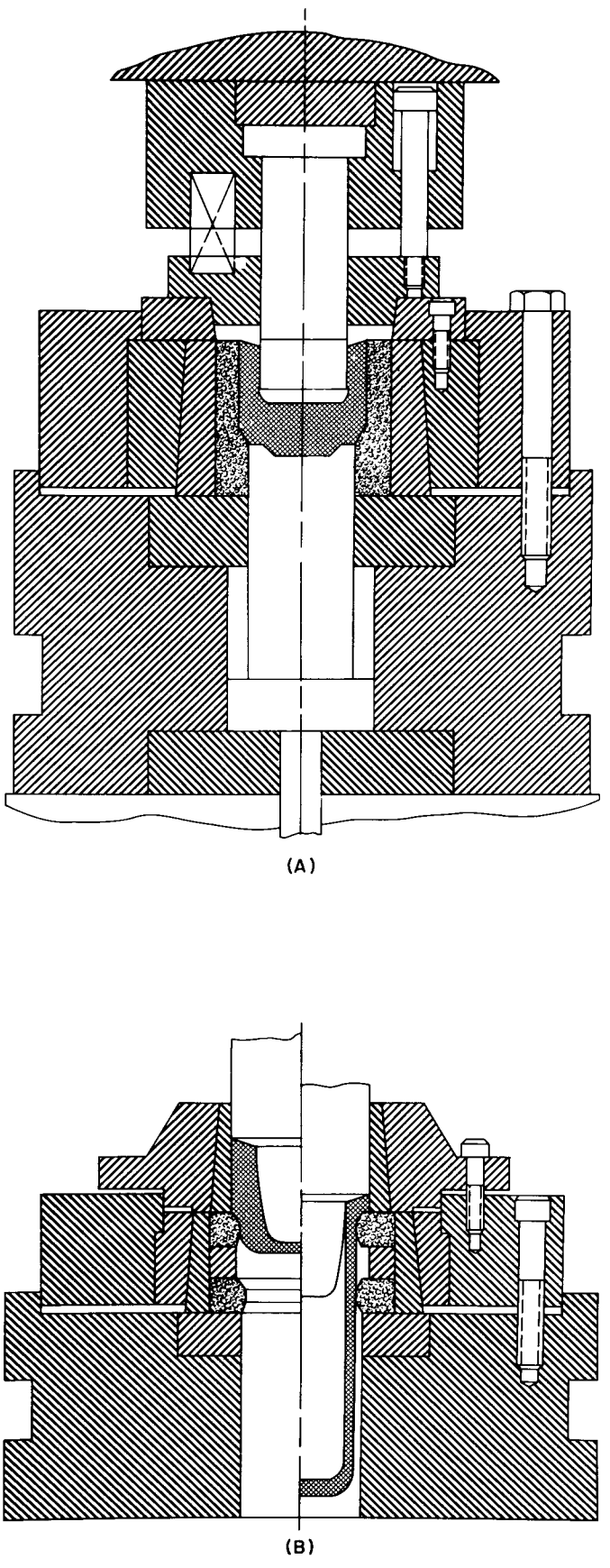


Figure 9. Extruding Dies: (A) backward, (B) forward

result, it is sometimes necessary to perform sizing operations in order to meet the design specifications.

TOOLING

A single-stage die is one in which only one operation, such as blanking, piercing, drawing, or redrawing, is performed.

A compound die is used to perform two or more operations with one stroke of the press. Operations such as blank and draw; blank and pierce; blank, draw, and pierce are typical examples of the operation performed on a compound die.

Progressive dies consist of a series of dies mounted on a single die set. Such a series may consist of a piercing station, a pierce and form station, and a notching station, or it may consist of a number of drawing and redrawing stations. With each stroke of the press, the part progresses through each phase of its manufacture until a completed part is ejected from the die.

Universal tools are interchangeable punches, dies, or draw rings which can be mounted on a die set.

Production schedules usually determine the type of die used in the manufacture of tube components. Single-stage and compound dies are used for the lower production schedules, while progressive dies are used for the higher schedules. Universal dies are used for the extremely low and short-run schedules.

The type of die used is also a matter of economics; single-stage dies cost less than compound dies, and compound dies are less expensive than progressive dies. If the capital expenditure of a tube type program is the determining factor in the production of a type, single-stage dies are generally built; but if the standard cost of the tube is more important, the more expensive compound or progressive dies are used.

Tool Costs

The cost of building tools depends upon the type of tool desired, as previously discussed, and the tolerances specified on the part drawing. The former is governed by production schedules and cannot be controlled by the designing or manufacturing groups. The design group, however, plays an important role in specifying the tolerances of the tube components.

Extremely close tolerances require equally close tolerances in the building of the tools — the closer the dimensional tolerances, the more costly the tool. Therefore, the tube designer must carefully consider tolerances and other specifications before standardizing the part. These tolerances should be as liberal as practical. Close tolerances should be specified only when absolutely necessary. Table I, by Watson Kintner, indicates the relationship between tolerances and cost for several operations.

Component Costs

Component costs depend upon three factors: material, labor, and "expense" (overhead). The material and expense, in many instances, cannot be controlled by tube designer. He must have a certain size and type of material for a given application; the expense of the manufacturing department is fixed by the financial structure of the organization. The labor in the manufacture of parts is, however, controllable to a great extent by the tube designer. Tolerances and other specifications, as well as designs, determine component costs. Additional operations may be required in order to meet the tube designer's specifications.

It cannot be stressed too frequently that additional operations such as sizing cost money, and to remain in a competitive market additional costs must be kept at a minimum. The tube designer must remember that one labor operation means an addition of from three to eight times that labor cost for expense alone. It can, therefore, be easily seen that tolerances should be as liberal as possible without affecting the operation of the tube or its manufacture, and that all critical specifications and designs for parts should always be discussed with the tool designer.

FOUR-SLIDE MACHINES

The four-slide machine, or multi-slide machine, is an automatic stamping or forming machine. As its name implies, this machine consists of four slides (front, rear, right-hand, and left-hand) mounted on the bed of the machine. Each slide is actuated by a cam. Fig. 10 illustrates a multi-slide machine.

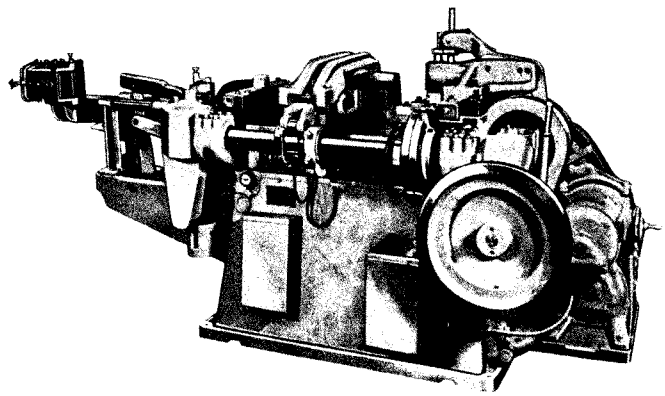


Figure 10. Multi-Slide Machine

In operation, the slides move in a horizontal plane toward each other and, with proper synchronization of the slides, each slide performs a function in the forming of material around a mandrel. The slides are moved by cams mounted on four shafts positioned around the machine. The power of the motor is transmitted to these shafts by means of a belt and clutch. A die-head assembly is mounted on the bed of the machine for such operations as notching, piercing, and embossing. The cut-off slide is mounted between the die-head assembly and the forming mandrel.

Multi-slide machines are used in the manufacture of such parts as plates, beam plates, various formed wire and strip parts, dynodes, and shields. Operations such

Table I

Relative Difficulty and Cost of Maintaining Tolerances for Various Forming Methods

	Machining		Stamping			
			Burr Not Involved		Burr Involved	
	Tolerance—% of diameter	% Tooling Cost & Maintenance	Tolerance—inch	% Tooling Cost & Maintenance	Tolerance—inch	% Tooling Cost & Maintenance
Easy	2.0	100	±0.004	100	±0.004	100
Difficult	1.0	200	±0.002	150	±0.002	200
Very Difficult	0.5	400	±0.001	200	±0.001	400
	Drawing*					
	4-Inch Diameter		2-Inch Diameter		1/2-Inch Diameter	
	Tolerance—inch	% Tooling Cost	Tolerance—inch	% Tooling Cost	Tolerance—inch	% Tooling Cost
Easy	±0.015	100	±0.015	100	±0.010	100
Difficult	±0.008	250	±0.007	250	±0.005	250
Very Difficult	±0.005	400	±0.003	400	±0.002	400
	Trimming**					
	4-Inch Diameter		2-Inch Diameter		1/2-Inch Diameter	
	Tolerance—inch	% Tooling Cost	Tolerance—inch	% Tooling Cost	Tolerance—inch	% Tooling Cost
Easy	±0.020	100	±0.015	100	±0.010	100
Difficult	±0.010	200	±0.008	200	±0.005	200
Very Difficult	±0.005	400	±0.004	400	±0.003	400
	Forming and Bending*					
	Molybdenum		Steel		Nickel	
	Tolerance—inch	% Tooling Cost	Tolerance—inch	% Tooling Cost	Tolerance—inch	% Tooling Cost
Easy	±0.030	100	±0.015	100	±0.012	100
Difficult	±0.015	200	±0.010	150	±0.008	150
Very Difficult	±0.010	500	±0.005	300	±0.004	300

*Difficulty depends on material and thickness.
**Difficulty depends on diameter.

as staking, lap seaming, lock seaming, swaging, and heading are also performed.

The simplest operation performed on the multi-slide machine is one in which material is cut to length. The material, strip or wire, is fed into the machine from the left, through a material straightener, and into the feeding mechanism. The feeding mechanism is adjustable, and any length, within the limitations of the machine, can be obtained by adjustment of the eccentric cam. The cut-off station consists of a die and a cut-off blade which is actuated by a cam mounted on the front shaft. The cut pieces drop through the bed of the machine into a pan.

In the forming of a part, the same operations as described above take place with the addition of the forming operations. For example, consider the manufacture of an L- or a U-shaped part, such as a brace or plate radiator. An instant before the material is cut from the strip, the front tool, which contains a holding pad, holds the material against the mandrel while the cut-off completes its function. The front tool continues its movement forward, bending the material around the mandrel. Similarly, for parts containing four bends, such as a rectangular band, the two side tools are employed to complete the forming of the U-shaped part. The side and rear tools are often used in overforming or sizing formed parts.

Beam-confining electrodes, or beam plates, required in power tubes, use a die head for the piercing, embossing, and notching operations. The die head is in essence a progressive die. A typical beam plate is made by embossing the part in the first station, piercing the front and rear windows in another station, and notching the upper and lower tabs in the third station. The notched strip is then formed in the conventional manner.

Many power-tube and receiving-tube plates are made by feeding two strips of material into the machine simultaneously with the notching performed on the rear strip, which is the wider of the two. The strips are separated by the mandrel and formed around it by the front and rear tools. A set of staking tools, which are a part of the forming tools, stakes both halves of the plate together. The staking tools, which consist of narrow lancing punches, lance both strips at one time, forming two sets of tabs and bending them away from the lanced hole. The staked plate is removed from the mandrel by a stripper which is actuated by an independent cam mounted on the rear shaft.

Parts such as bead supports are swaged and headed on four-slide machines; the front and rear tools swage the paddle and the right-hand tool forms the head.

Countless parts can be made on these machines but, because of the expense involved in the tooling and set-up, and the speed with which the machine is capable of running, production is limited to high-volume parts. Four-slide machines produce plates at a rate of 5,000 parts per hour, and have speeds as high as 25,000 parts per hour for bead supports.

Processes Used in Metal Parts Fabrication and Finishing

H. A. Kauffman

Lancaster

This chapter is written with two main considerations: first, to illustrate the effect of parts design on processes and costs, and second, to describe process factors which are peculiar to electron-tube manufacturing.

The subjects covered include resistance welding, electroplating, chemical cleaning of parts, electropolishing, and photomechanical etching.

Generally known and published information is included only where it is necessary for understanding of the topic being discussed. It is suggested that standard references be used if additional information is desired.

RESISTANCE WELDING OF ELECTRON TUBE PARTS AND ASSEMBLIES

The resistance welding of electron tube parts and assemblies is a delicate process requiring precision and quality workmanship. The general rules presented in welding reference books do not provide for the special requirements of electron-tube manufacturing. If the design of parts can comply with the special considerations necessary for good welding, quality can be improved, and costs can be reduced.

Standard resistance-welding equipment comprises three different systems:

1. A mechanical system that aligns the parts and applies the required welding pressure.
2. An electrical circuit that conducts the regulated current through the electrodes.
3. A control system that times the current and regulates the pressure.

It is the mechanical system of the welding equipment to which the part must be adapted. Alignment fixtures and electrodes are designed to meet the part requirements. Therefore, the part design should be such as to utilize the best possible welding methods for the particular requirements.

The following weld requirements should be evaluated for all parts:

1. Position and size of weld
2. Strength of weld
3. Amount of weld splash
4. Alignment of parts
5. Surface condition of assembly

If at all possible, these requirements should be stated in terms of physical dimensions and not on a basis requiring human judgement. The use of physical dimensions is the only satisfactory way to insure proper quality control.

Three general methods are used to weld electron tube parts: spot welding, projection welding, and seam welding. The special problems in meeting the five requirements with these three methods will be discussed.

Spot Welding

Spot welding is the simplest of the three methods, although it is the slowest in production and the hardest to control. Consider the five general weld requirements:

Requirement 1. The position and size of weld is determined by the position, size, and shape of the electrode used in welding.

Requirement 2. The strength of the weld, although a function of the heat generated, is limited by other factors. The larger the size of the weld, the higher the strength, but there must be sufficient thickness of material to prevent "burning" through. Over-all strength of the assembly can be improved by critically placing each weld to counteract the forces to which the assembly will be subjected.

Another important factor is that of adjacent welds. Welds made too close together are robbed of their strength because the welding current flows back through the previous welds, instead of forming a strong new weld. In other words, five welds per inch could be stronger than ten.

Requirement 3. Weld splash, because of the nature of electron tubes, is very undesirable. With a given part design and strength requirement, splash is the most difficult to control of all the weld requirements. If full consideration is given in the part design, weld splash problems can be reduced. In general, the greater the strength, the greater the weld splash. The part thickness should be sufficient to obtain adequate strength

without approaching the point where splash occurs. In many cases, a 0.010-inch thickness material is expected to weld as rigid as if it were 0.020 inch. Because of the great number of part sizes and shapes, it is impossible here to provide figures that will predict when splash will occur. Therefore, engineering thinking plus the use of tables in the Welding Handbook[1] will have to be applied in each case. One rule that will help in the prevention of splash is to locate the weld in from the edge of the material, thus allowing some splash to be absorbed between the two parts. The condition of the welding surfaces also affects the amount of weld splash. The smoother and cleaner the surface, the less weld splash there will be. Foreign material or high points create a very high current density causing the metal temperature to reach the melting point so that splash occurs. The heating of occluded gases in the metals being welded also results in the splash of molten metal.

Requirement 4. Alignment of parts must be done by some mechanical means. The welding electrodes generally have no effect on parts alignment. If alignment is critical, a specific reference surface should be provided in the parts specification.

Requirement 5. The surface condition of the assembly in some cases is important. Two surface problems can occur: (1) if a small part is not provided with an adequate method of removing excess heat, oxidation can occur; (2) if the resistance between the electrode and part is high, pitting, copper deposit, or sticking will occur. These surface effects depend mostly on the materials used for the parts and the electrodes and on the welding pressure. Tables in the Welding Handbook[1] can help in the selection of suitable parts materials.

Precautions for a typical spot welding operation are:

1. Provide flat clean surfaces for minimum temperature at points 1 and 2 of Fig. 1.

2. Allow clearance in part for electrode to prevent electrical shorting to part.

3. Allow sufficient part thickness for required strength

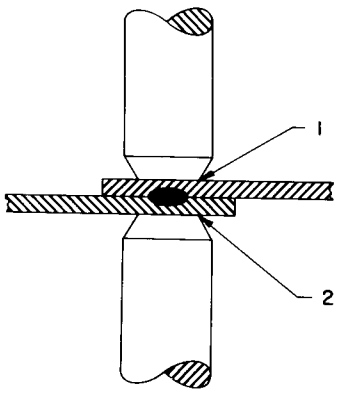


Figure 1. Typical Spot-Welding Operation

Projection Welding

Designing parts for projection welding is more difficult than for the other welding methods. However, projection welding has the greatest advantage, that of being capable of giving consistent welds in high production. All of the requirements listed and discussed for spot welding apply to projection welding plus the following:

Requirement 1. The position and size of weld is determined by the position and size of the projection; thus, the control is in the design of the part. For high production, this method of control is a primary advantage.

Requirement 2. The strength of the weld will be dedetermined by the size and shape of the projection. A small projection will produce a weak weld. Because the pressure may build up to maximum before current is applied, the projection is frequently crushed before current flows. The value of the projection is lost; current density is low, and a weak weld results. Properly designed projections produce stronger welds by permitting sufficient pressure to be applied to bring the surfaces into intimate contact, thus allowing the use of higher currents and reducing porosity and internal cracking of welded metals.

Requirement 3. The amount of weld splash is the greatest disadvantage of projection welding if the projection is poorly designed. The correct projection will make strong consistent welds relatively free from weld splash. The important factors are: (1) the contact area between the parts and the electrode should be evenly distributed and should be considerably greater then the area of the base of the projection; (2) the surface of the projection should be absolutely uniform on any one part. If there is more electrode contact on one side of a projection than on the other, that side will have a greater rate of electrode wear and, therefore, the pressure across the areas of contact will become nonuniform. A uniform current density across the weld is absolutely necessary for the prevention of weld splash.

Requirement 4. Alignment of parts can best be obtained by use of separate mechanical means other than by the use of the shape or position of the electrodes. When electrodes are used for part alignment, electrode wear is considerable, and therefore, the alignment of the electrodes cannot be maintained.

Fig. 2 illustrates the effects of different circular bosses for projection welding. Fig. 2A illustrates poor design having the following faults: (1) point contact to electrode, (2) weak embossing, (3) rough surface, (4) loose height tolerance. Fig. 2B illustrates good design having the following features: (1) area contact to electrode, (2) strong embossing, (3) smooth surface, (4) tight height tolerance.

Seam Welding

Designing parts for seam welding is almost as simple as for spot welding. However, seam welding presents the greatest difficulty of all the methods in obtaining consistently uniform welds. Its advantage lies in the high rate of production that can be obtained for a par-

ticular requirement, when the quantities required are sufficient to justify the more elaborate equipment and setup needed for this method.

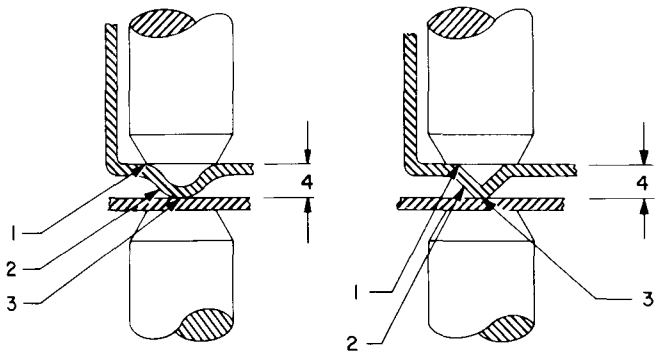


Figure 2. Typical Projection-Welding Operation

All of the requirements listed and discussed for spot welding also apply for seam welding plus the following:

Requirements 1 and 4. The position of the weld and the alignment of parts are interrelated. In this method of welding, the parts are continuously moving during the welded operation. The parts must be aligned and then accurately positioned under the roller electrodes. Where alignment is very critical and difficult to achieve while the parts are in motion, the parts may be "tacked" together by a few small spot welds before the seam welding.

Requirement 3. Weld splash in seam welding usually occurs at the ends of the weld. Therefore, it is important that the weld should not start nor end close to the edge of the material.

Requirement 5. The surface condition of the assembly sometimes becomes a problem because discoloration occurs. The cause is the increased heat due to the continuous flow of current.

Precautions for a typical seam-welding operation are:

1. Center seam weld on welding land and terminate before edge
2. Tack weld to align parts before seam welding
3. Provide adequate welding land width

ELECTROPOLISHING

Electropolishing is a process of removing metal electrolytically by using direct current in an acid solution. In essence, electropolishing is a deplating operation where the work is anodic, thus effecting the removal of metal from high spots and ragged edges; the low spots are fairly passive and little metal is removed from them.

Electropolishing produces a brilliant lustrous finish. Parts or shapes that are intricate and would be difficult to polish by chemical or mechanical means may be polished easily by electropolishing. A rough surface can be smoothed out to eliminate undesirable high spots and to obtain a more uniform surface.

Ragged or burred edges may be made smooth; however, the process has limitations for this application. If excessive burr is present, it is practically impossible to remove the burr electrolytically and still maintain close tolerances. The extreme amount of electropolishing needed to remove the burr will also affect the other metal edges. In electropolishing kinescope apertures, severe polishing leads to deviation in concentricity of the hole and also causes excessive enlargement of the hole.

While electropolishing can be a very useful process, it has some disadvantages. This process can uncover surface defects in metals and, thus, produce a rough surface rather than a smooth one. Severe polishing can also expose nonmetallic inclusions and surface impurities. Best polishing results are obtained by careful specification of the physical properties of the metal to be polished.

In general, electropolishing is much cheaper than mechanical buffing, but is more expensive than tumbling or chemical bright dipping.

CHEMICAL BRIGHT DIPPING

Chemical bright dipping is a selective pickling process which removes oxides but has limited effect on the base metal. The desired controlled action is usually obtained by the addition of inhibitors to the solution.

The chief advantage of this process is that no electrical contact or current is required, thus permitting the rapid handling of quantities of small parts. The reduced handling reduces labor costs. Parts which are shaped so that they nest or are difficult to rinse are very difficult to process.

The main disadvantage of bright dipping is the problem of achieving control and uniformity. Many bright-dipping solutions necessitate very critical control conditions. Failure in control usually results in overetching or nonuniformity or both. Continued bright dipping to obtain uniform appearance may result in nonuniformity of critical dimensions.

In general, the cost of bright dipping is less than that of mechanical buffing or electropolishing, but is greater than that of tumbling.

PHOTOMECHANICAL ETCHING

The term photomechanical etching refers to the manufacture of a precision metal part from relatively thin (up to 0.010 inch) sheet metal through the use of photographic techniques and chemical etchants rather than the more usual mechanical methods. This process is, in essence, a reduction of the photoengraver's art to mass-production techniques. Through proper choice of etchants, the process may be applied to many metals such as zinc, copper, and steel. The process is characterized by very good reproducibility and by a high degree of precision; photomechanical etching is capable of meeting tolerances of 0.001 inch or less.

The process consists of four main steps:

1. Sensitizing the metal sheet with a photosensitive coating material
2. Printing by exposure to intense light through a photographic pattern
3. Developing of the photographic image
4. Chemical etching of the part which duplicates in metal the image previously produced in the coating

Photomechanical etching can be used for very simple parts (such as could be produced on a punch press) and at the same time is adaptable to such complicated designs as the shadow mask used in the color kinescope. This mask is a sheet-metal disk containing several hundred thousand holes all of which must meet tolerances of less than 0.001 inch in location, size, and shape.

Sensitizing

Before the application of the photosensitive coating, the metal must be made scrupulously clean and free of grease or oxides. Grease may be removed by any of the standard solvent, emulsion, or caustic cleaning techniques, although a spray-type caustic-cleaning unit is probably one of the most satisfactory for the processing of continuous strip material. After the metal is thoroughly rinsed, oxide may be removed by spray-pickling or acid-washing. This acid washing serves a secondary purpose in that it produces a uniform matte surface which provides a good base for adherence of the coating material and aids in reducing photographic exposure time in subsequent processing. All traces of the acid are rinsed off the metal before processing it further.

The metal strip is then coated. A number of coating materials may be used; a water solution of fish glue made light sensitive by the addition of ammonium dichromate is one of those more commonly used. Some care must be exercised since the materials are light sensitive in solution as well as after application. The coating material is usually applied by flowing the liquid coating vertically down the metal strip from one or more flow nozzles. At this stage, great care must be exercised to prevent contamination of the coating material or the still-wet, coated metal with small particles of dirt which may produce serious defects in the finished part through the formation of holes in the coating. After coating, the coated strip is dried by the application of mild heat. Excessive heat must be avoided since the majority of the coating materials in use are sensitive to heat as well as to light. Temperatures of 150 to 225 F are normally satisfactory.

Printing

The pattern is produced in the coating by exposing it to intense light through a photographic negative pattern. The light sources commonly in use are those rich in blue and ultraviolet light such as carbon- or mercury-arc lamps. Exposure time and intensity will vary with the coating thickness and with the surface condition of the base metal. An exposure of 11,500 foot-candles for three minutes has been found satisfactory for a dichromated glue coating of 0.00045-inch thickness.

Depending on the thickness of the metal used and the nature of the part involved, the pattern may be produced on one side only or on both sides of the coated metal simultaneously. If the pattern is to be produced on both sides, mirror-image negative patterns must be held directly opposite one another. Perfect alignment on opposite sides of the coated metal is obtained through the use of line up "cross hatch patterns" or "registry" marks printed on both negative patterns. The negative patterns are normally produced by photographing the outline and features of the desired part on film or on a glass plate. This photograph may be made from the part itself, another photograph, or a precision drawing.

The exposure is made in a photographic contact printing frame in which the photographic pattern and the coated metal are held in intimate contact through the use of a vacuum-seal arrangement and vacuum pumps. A vacuum of 25 to 29 inches of mercury is required for precise work. Proper contact of the metal and the photographic pattern may be determined by observation of the Newton ring pattern formed upon application of the vacuum. The image produced upon exposure will be similar to a standard photographic-emulsion latent image with this exception: The portions of the coating exposed to the full rays of the light source become insoluble in water (or other solvents in the case of coating other than dichromated glue). Those portions which are masked by dark portions of the photographic negative are unexposed and remain soluble.

Developing

The latent image is developed by thoroughly washing the metal strip with water (or other solvents) to remove the still-soluble portions of the coating. All unexposed coating must be removed or precision will be lost. A wetting agent is added to the final rinse to reduce surface tension and promote uniform drying. The strip is dried in a high-temperature oven (about 550 to 650 F); during this process, the coating is also partially carbonized to form a glassy acid-resistant coating. Proper temperature is essential since underbaking does not produce satisfactory acid resistance, while overbaking promotes complete carbonization and peeling of the coating. When baking is completed, a coated strip with a photographically produced pattern is obtained in which coating remains where metal is to remain in the finished part, and in which the coating has been removed where metal is to be removed.

Etching

The final operation, etching, is again preceded by a cleaning operation utilizing such agents as acids, scouring powders, brushes, etc., in order to remove the last traces of undesired coating and oxides produced in the previous baking operation. After cleaning and rinsing, the processed metal strip is passed through the etching chamber itself. There the etching acid is splashed or sprayed on the coated metal strip. In the case of steel, a slightly acidified (0.1 per cent free HCl) 44° Baume solution of ferric chloride is used at temperatures from 70 to 150 F; the higher temperatures result in increased speeds. More-dilute solutions result in increased speeds, but will also result in a rough, granular etched surface in the case of steel.

Etching produces a part in which uncoated areas have been dissolved by the etching acid while those areas still protected by the coating remain untouched. The part may be allowed to fall from the metal strip at this point or may, as is usually done, be permitted to be retained in continuous strip form, and later removed in a manner similar to tearing postage stamps from a perforated sheet. After again rinsing, the remaining coating is removed by flooding or spraying the metal with a caustic solution of the paint-stripper type maintained at elevated temperatures. After coating removal, the metal is rinsed again, and may again be cleaned using acids and brushes if desired. After final water rinsing, the metal strip is dried and the parts removed from the metal strip.

ELECTROPLATING

Electroplating is the deposition of positive metallic ions from an appropriate solution on to the piece to be plated. The piece is attached to the negative terminal of a dc power supply. An anode must also be suspended in the solution and be connected to the positive terminal of the power supply. Anodes may be made of either soluble or nonsoluble material. Examples of nonsoluble anode materials are lead in chromium plating, carbon in nickel plating, and steel in cyanide copper, silver, and gold plating. Anodes of soluble materials are more frequently used because their dissolving tends to maintain a uniform level of metallic ions available for deposition from the plating solution. Of course, in this case the anode material is the metal being plated.

The more commonly plated metals are listed in Table I. Included in the table are the common current densities in amperes per square foot, the time for the deposition of 0.001 inch at the given current density, and the more common uses of each type of plating in electron-tube manufacturing.

Table I

Metal	Current—amp/ft^2	Minutes to Plate 0.001-in.	Reasons for Use
Nickel	40	31	appearance, corrosion resistance, brazing surface for silver soldering
Copper	20	26	appearance, conductivity
Silver	5	74	appearance, conductivity
Gold	5	45	appearance, emission reduction, conductivity, corrosion resistance
Chromium	288	64	appearance, sealing, wear resistance

As may be observed from the table, all plating metals may be applied for improving appearance. When appearance is the main purpose for plating, special bright-plating solutions designed to produce maximum eye-appeal with a minimum plated thickness are used.

Nickel plating is specified primarily for its corrosion resistance but is frequently used for electron tube parts which are assembled by brazing with silver solder. In the latter application the nickel prevents silver-solder penetration of Kovar thus eliminating a source of vacuum leakage.

Copper, silver, and gold are used to improve surface conductivity, especially at high frequencies. Other factors usually determine which of the three metals is specified. Gold has the advantage of little corrosion compared to the rapid tarnishing rate of silver and the somewhat slower rate of copper oxidation. However, gold is extremely costly and should be used only where very definite quality advantages exist.

Other less common plating applications are the use of gold to reduce secondary emission and the use of chromium for wear resistance and for high-conductivity glass seals.

Effect of Parts Design

The size of a part frequently cannot be altered, but is determined by the tube requirements. However, it is well to remember that when plating is used the parts must be handled through a half-dozen or more process steps. The weight to be lifted by an operator, the jigging or holding of the part and the high current requirements are important factors that must be considered. When large size is involved, the plating of smaller subassembly parts may be cheaper and more efficient. With small parts, the plating of assemblies is usually less costly.

The shape of parts is much more critical to satisfactory plating than is the size. Numerous holes are desirable to permit easy jigging and escape of trapped gases and air during the plating and treatment cycles. The entrapment of air or gas will not only cause poor preparation and cleaning but will prevent plating of surfaces not in contact with plating solutions. While cup-shaped parts can be racked to prevent the formation of an air pocket, problems arise because solutions are carried over from one process to another.

Recesses and internal areas are difficult to plate and often require special conforming anodes or internal anodes. Irregularities in surfaces produce nonuniform plating thicknesses. Edges and outside areas will generally have heavier plating than recesses and internal areas. Designing for these thickness variations will help to reduce costs when general parts design cannot be simplified.

Metal areas which are insulated from each other necessitate special racking for electrical contact. Here again proper design can greatly simplify the racking problem and, thus, reduce the handling costs of plating.

Requirements for Nonplated Areas

The specification of nonplated areas on a part or assembly cause plating complications and increase cost. Nonplated areas are usually obtained by masking the nonplated area with lacquer or masking tape, or by the

use of special jigs and fixtures. The cost is directly proportional to the dimensional tolerances of the masked area and to the difficulty in masking the area. Inside diameters are usually more difficult to mask than outside diameters because the masking material tends to shrink and pull away from inside surfaces. Thin or "hairline" areas are difficult to produce and maintain during processing.

Even the best designed jigs and the best applied masking will frequently allow edge seepage of solution and permit plating where it was not desired. The need for sharp definition of the plated and nonplated areas should be avoided wherever possible.

Another method for obtaining unplated areas is to strip the plating after it has been applied. This method is generally not recommended because of the danger of attack on the base metal. The method is time consuming and difficult to control when clear-cut boundary lines are to be produced.

Requirements for No Nonplated Areas

In contrast to the difficulty of maintaining nonplated areas, is the equal difficulty of obtaining completely plated areas. Wherever wire or rack contacts are made to the parts, bare or thin spots result. These spots can be eliminated by moving the parts during a single plating cycle or by successive plating applications. The extra handling leads to higher labor costs and increases the possibility of rejects due to the interrupted plating cycle.

Large parts may require several hundred amperes of plating current. With such parts, electrical contact and proper current distribution through the part must receive careful attention.

A frequent cause of nonplated areas is the alloying of the plated metal with the base metal or other previously applied metal. These "bare" spots are usually seen only after a high-temperature firing treatment and should not be considered cause for rejection unless tube application is definitely affected.

Cleaning for Plating

Methods for cleaning parts to be plated can be divided into four general categories: degreasing, firing, acid pickling, and mechanical cleaning. Common degreasing methods involve soaking or agitation in special solvents or alkaline solutions. Agitation may be supplied manually, mechanically, ultrasonically, or by air or electrolytic means described in more detail later. Solvents, such as, trichlorethylene, acetone, and alkaline compounds, are commercially available, usually with generous amounts of engineering service from the product vendors.

Electrocleaning is usually applied in all plating processes in addition to all other cleaning operations. Electrocleaning solutions are generally alkaline, but for specific application may be acid solutions. The part is made anodic or cathodic with an accompanying evolution of gas. This evolution of gas produces a scrubbing action thus giving very efficient film removal on part surfaces. Cathodic cleaning is more effective because of the greater volume of hydrogen gas evolved. Preference for anodic cleaning is usually designated when there is danger of hydrogen penetration or when cleaning solutions may contain and deposit some metallic ions.

For electron-tube parts, degreasing of the parts followed by a water wash is frequently followed by firing the parts at high temperatures, 900 to 1100 C, in dry hydrogen or line hydrogen. This firing can produce oxide-free surfaces which often can be plated with no additional chemical treatment. If firing is required for degassing or some other reason, it may be well to fire just prior to plating as an aid to better cleanliness.

The third cleaning method is acid pickling. Pickling is used primarily to remove oxide. Greatest problems occur in pickling from the combinations of metals used in assemblies. Satisfactory acid solutions for copper cleaning will not do as well for iron-base metals. In the more difficult pickling jobs, combinations of pickling and mechanical cleaning methods are often used.

Sandblasting and wire brushing are the more common types of mechanical cleaning. There are many variations of these methods from hydroblasting to the use of steel wool; each application should be fitted to the complexity of the part or assembly being cleaned.

Striking

Striking is the application of a thin coat of electrodeposited metal to the part. The striking is usually done in a bath with high acid or high cyanide and low metal content. There are three main reasons for striking. The first is to prevent immersion plating which is particularly prevalent with copper and silver plating. Immersion plating is the chemical replacement of a metal by one with a greater electropotential. Immersion plating is undesirable because of its generally poor adherence. The second reason is to activate the surface of the base metal while the initial layer of plated metal is deposited. Activation is the removal of a passive oxide film which has not been removed by prior cleaning operations. Since such films are characteristic of nickel and nickel alloys, activation is usually necessary when these metals are to be plated. The third reason for striking is to obtain quick, uniform coverage of different base metals for the subsequent slower application of the required thickness of plating.

Plating Requirements and Costs

The more preparation that is required for satisfactory plating, the greater the cost. Care in handling parts and assemblies before plating can reduce the preparation required. Be aware that clean surfaces are required for plating. Keep parts and assemblies simple and easy to clean. Avoid contaminants, during brazing or other assembly operations, which produce cleaning problems.

Keep plating thickness requirements to a minimum not only for cost purposes but for quality. Heavy plating may develop internal stresses which may cause cracking and peeling. Thin plating will assist in allowing gas

diffusion through the deposit during firing or in later tube processing thus eliminating the development of blistered plating.

As mentioned earlier, specify controlled thickness only in critical areas. Do not add restrictions on edge build-up or on thin areas in recesses unless absolutely necessary.

In addition to the cost factors which have been mentioned, handling methods involve cost considerations. Small-volume plating requirements are usually met by racking the parts individually or in small groups and carrying the racks by hand through the various cleaning and plating cycles. Where volume is high, automated or conveyor handling of parts can be used. Conveyor handling reduces labor costs, but greatly increases equipment costs. A common mass-production method for small parts is the use of plating barrels. This method is especially practical where entire parts can be plated and when parts will tumble freely and not nest in one another.

A special type of plating, periodic reverse (P. R.) plating, is seldom used because of its high cost. In P. R. plating, the current is actually reversed periodically during the plating cycle. The advantage of this type of plating is the leveling effect, produced on the deposit, which tends to eliminate the high and low spots. This method, with careful adjustment, can produce special-finish appearances.

Treatment After Plating

The first major requirement after plating is an adequate rinsing. If parts have been masked, masking materials must be removed and the parts must be thoroughly rerinsed. Special cleaning of unplated areas may be required to insure removal of all undesired plating. Drying becomes expensive where water spotting is objectionable. Special solvents to remove water from parts and to reduce staining, along with individual drying of parts, greatly increases finishing costs. Where cleanliness rather than freedom from spotting is the criterion, care must be taken to avoid the excessive treatment required to produce spot-free parts.

Inspection of Plating

Plating adherence is generally tested by firing at temperatures of 800 to 1000 C. In addition to testing adherence, this firing operation causes alloying at the interface of the plated and base metals. Also, in the case of gold-plating molybdenum wire for grids, this firing appears to remove base metal oxides and to permit the bonding of gold to the molybdenum.

Plating thickness can be measured by making metallurgical cross-sections of representative samples. This method, of course, is destructive but where tight control is essential, it is a must. Several chemical tests, also destructive, are in common use. These tests are less accurate but can be more rapidly performed than the cross-section method. All of these chemical tests depend on measurement of the time for the removal of the plated deposit from a controlled area on a part. An anodic solution method, a controlled deplating, is the most accurate of the numerous chemical methods.

Nondestructive thickness measurements are sometimes made by weighing the deposited metal and determining the area covered by the plating. The resultant thickness can then be calculated by dividing the weight by the product of the area and a constant which varies with the metal plated. Generally, more accurate and rapid nondestructive measurements can be made with commercial test devices that depend on either magnetic or eddy-current measurements. These magnetic or eddy-current measurements are expensive in terms of time for initial calibration and setup, but are inexpensive and rapid for daily production measurement and control tests.

Restrictive plating specifications in conjunction with complex part design have generally resulted in sufficient plating variation to require 100 per cent visual inspection of all parts. When such inspection can be avoided and a sample inspection performed, cost can be reduced.

Summary

In summary, when possible consult a plating expert before designs are completely developed. Whenever possible, review some of the problems existing on current production types. Low plating cost can be a design requirement along with utility and appearance requirements.

REFERENCES

1. The Welding Handbook, American Welding Society, New York, 1955
2. The Welding Encyclopedia, Welding Engineering Publishing Company, Chicago, 1943
3. Resistance Welding, Theory and Use, American Welding Society, Reinhold Publishing Corporation, New York, 1956
4. Metal Finishing Guidebook Directory, Finishing Publications, Inc., Westwood, 1960
5. Gray, Allen G., Modern Electroplating, John Wiley and Sons, Inc., New York, 1953
6. Blum and Hogaboom, Principles of Electroplating and Electroforming, McGraw-Hill Book Company, New York, 1949

Glass for Receiving Tubes

J. Gallup

Harrison

DEFINITION

Glass is a homogeneous material made by the fusion of various inorganic materials, mainly oxides, into a uniform solution which is then undercooled to approximate the solid state but without the formation of crystals. Thus, at room temperature, it exists as an undercooled liquid, the viscosity of which is so great that for most practical purposes the glass is a noncrystalline solid.

The glasses used in electron tubes are divided, in the main, into two broad classes termed "hard" and "soft." These terms refer to the melting temperatures required in making the glasses, and have little direct relation to the scratch hardness or abrasion resistance of the glasses. The higher the melting temperature, the "harder" the glass is said to be. Conversely, the lower the melting temperature, the "softer" the glass. The hard glasses used in electron tube manufacture require working temperatures above 1000 C, whereas the soft glasses can be worked or formed at temperatures below 1000 C.

Receiving tubes have been made mostly from soft glasses, with only a few special-purpose receiving tubes using the more refractory hard glasses. Soft glasses have been preferred because of the greater cost of making and working hard glasses, and because the low thermal expansions of hard glasses require relatively expensive lead-in or sealing materials, such as tungsten, molybdenum, or Kovar. On the other hand, soft glasses, having higher thermal expansion, can be used with cheaper sealing metals, such as dumet, chrome iron, and iron-nickel alloys. It seems likely, however, that the growing need for higher power outputs and consequent higher operating temperatures will increase the use of hard glasses in receiving tubes because of the greater temperature stability of such glasses.

A third class of glasses known as "solder glasses"* has come into limited use in recent years. These glasses are "softer" than conventional "soft" glasses and have working temperatures below 700 C. Solder glasses are used, as their name implies, to "solder" or seal other materials together at relatively low temperatures (400 to 700 C). Because solder glasses are weaker and less resistant to weathering or attack by water that most other glasses, they are used only where glasses having higher working temperatures cannot be used without oxidizing or otherwise damaging contiguous mount-structure parts.

COMPOSITION

Glass is composed of two types of constituents: network (or glass) formers, and network modifiers. Silicon dioxide is the most common commercial glass former. The continuous random networks of silicon-oxygen tetrahedra having silicon atoms at the center have such bonding characteristics that they exhibit high viscosities even at temperatures far above the melting point of the crystalline compound silicon dioxide (SiO_2). This high viscosity resists forces tending to cause crystallization and, hence, retards crystallization strongly. Other network formers in common use are boron trioxide, phosphorus pentoxide, and vanadium pentoxide. Network modifiers are materials, such as the alkalies and alkaline earths, which do not form glasses by themselves because of their strong tendency to crystallize or, alternatively, because they cannot readily form random network structures. However, if the addition of network modifiers is kept below certain critical limits, they can enter into and modify the random network structures of the glass formers without producing crystallization. Sodium, potassium, lithium, calcium, and lead oxides are some of the network modifiers commonly used in commercial glasses to obtain the proper combination of desirable characteristics which is not available in any glasses made solely of glass formers. Such desirable characteristics vary with the use of the glass; however, in receiving-tube stem glasses, they are low working temperature, high thermal expansion, high electrical resistance, and insolubility.

The chemical compositions of four typical commercial glasses are shown in Table I. It should be noted that all contain approximately 60 per cent or more of silicon dioxide. Pure silica glass, called fused quartz, contains 100 per cent silicon dioxide and is the most stable glass normally available. Unfortunately, its high melting range (approximately 1700 degrees Centigrade), and its low thermal expansion (0.5×10^{-6} inches per inch per deg. C) render it difficult to work and hard to match to the thermal expansion of metals. Therefore,

*Gallup, J. and A. G. F. Dingwall, "Properties of Some Low-Temperature Solder Glasses," Bull. Am. Cer. Soc., 36 (2) pp. 47-51, 1957

the desirable properties of silica glass are taken advantage of by its use as the major constituent in electron-tube glass, and its undesirable properties are modified by the addition of generally lower-melting, higher-expansion, network modifiers. In some cases, however, the network modifiers may be higher-melting oxides, such as calcium oxide and magnesium oxide.

Table I

Glass Compositions in Weight Per Cent of Oxides*

Oxide	Type of Glass			
	Pyrex	Lime Window	Lime Bulb	Lead Bulb
SiO_2	80.75	72.14	71.5-73.5	58-59
B_2O_3	12.00	---	---	---
Al_2O_3	2.2	1.06	1-2	0.2-0.4
Fe_2O_3		0.15	0.06-0.1	0.04-0.06
CaO	0.30	11.24	5-6	---
Na_2O	4.10	12.6	14-17	7-8
K_2O	0.10	---	0-1.5	4-5
As_2O_5	0.40	0.01	---	---
PbO	---	---	---	28-30
MgO	---	2.62	3.5-4.5	---

*Sharp, Donald E. "Chemical Compositions of Commercial Glasses," Industrial and Engineering Chemistry, 25(7), pp. 755-64, 1933

PRINCIPAL PROPERTIES OF GLASS

The properties of glass most important to its use in electron tubes are thermal expansion, viscosity, and electrical resistivity.

Thermal Expansion

Thermal expansion is important because it controls the ability of the glass to stay sealed to another substance without fracturing when cooled from the sealing temperature. It is measured in units per unit of length per degree change of temperature. In the properties shown in Table II, the coefficient of thermal expansion is expressed in inches per inch per degree Centigrade. In practice, a difference of 1.0×10^{-6} in the coefficients of expansion is about the greatest that can exist between two different glasses sealed to each other without danger of rupture occurring when the seal is cooled from the sealing temperatures. The thermal-expansion coefficient is determined from the essentially straight-line portion of the thermal expansion curve between room temperature and 300 C, and compatible pairs are selected on this basis in one method of expansion matching.

Fig. 1 shows that the curves start to bend upward at higher temperatures near the strain point, in some cases so sharply that the coefficient of expansion increases to as much as three times the low-range value. Fortunately, the glass is in the annealing, or slow-yielding, range when this rapid increase occurs, so that large strains can be avoided by the correct control of the cooling process. Otherwise, it would be very difficult to match different glasses for thermal expansion.

A more refined method of matching thermal expansion makes use of a value called the "effective setting point" which, for normal industrial cooling rates, is about midway between the annealing and strain points. At the effective setting point, the glass has a viscosity of 10^{14} poises. The effective setting points are indicated on Fig. 1 by vertical markers. When this matching method is used, successful seals are not considered possible when the $\Delta L/L \times 10^3$ difference between the setting points is greater than one.

The thermal expansion coefficients (the slopes of the straight-line portion of the curves of Fig. 1) of normal soft glasses cover the range from approximately 8.0 to 13.0×10^{-6}. These coefficients for hard glass encompass the range from about 0.5 to 5.0×10^{-6}.

Some common glass-to-metal seal materials are listed in Table III.

Although all the glasses in any one group will seal to each other, they usually require special intermediate sealing glasses to match them to glasses of other groups.

Copper can be sealed to any of the glasses of any of the groups by use of the Housekeeper seal principle, by which the glass is sealed to a thinned or flattened edge of the copper. This type of seal depends on the ability of the thin copper (0.010 inch or less) to flow or yield so that the stress in the glass remains below its elastic limit or rupture strength.

Viscosity

In theory and in practice, viscosity is the property that distinguishes glass from crystalline solids. It is the property, common to all liquids, of resistance to flow under the influence of gravity or other forces. In theory, glass is always a liquid and, hence, is always viscous. In practice, its long viscosity range allows glass to be worked, formed, and annealed without crystallization. Viscosity is measured in units called poises. A liquid has a viscosity of one poise when it separates two parallel planes one centimeter apart, and a tangential force of one dyne per square centimeter applied to one plane will cause a movement of one centimeter per second of this plane with respect to the fixed plane. The actual magnitude of the unit can be realized when it is recalled that water has a viscosity of 0.01 poise at 20.2 deg C. One experimenter has calculated by extrapolation form measurements made at higher temperatures that glass should have a viscosity in the neighborhood of 10^{78} at room temperature — a value which represents an enormous resistance to flow — far above the elastic limit or rupture strength of the glass.

Table II

Physical Properties of Some Representative Glasses

Code No.	Lab. No.	Type	Density g/cc	Coefficient of Expansion in/in/ C	Working Point C	Softness C	Annealing Point C	Strain Point C	Refractive Index	Specific Resistance — ohm-cm	
										at 250 C	at 350 C
SOFT											
0010	G1	Potash-soda-lead exhaust tubing	2.85	9.1×10^{-6}	970	626	430	395	1.542	$1{,}190\times10^{6}$	13.3×10^{6}
0080	G8	Lime bulb	2.47	9.2×10^{-6}	1000	696	510	470	1.512	2.27×10^{6}	0.148×10^{6}
0120	G12	Potash-soda-lead stem	3.05	8.9×10^{-6}	975	630	435	395	1.557	$11{,}900\times10^{6}$	95.5×10^{6}
8160	G814KW	Potash-lead stem	2.98	9.1×10^{-6}	975	627	435	399	1.553	$36{,}300\times10^{6}$	237×10^{6}
1991	G184ET	Iron sealing	3.18	12.7×10^{-6}	---	539	393	366	---	---	21.6×10^{6}
8161	814UD	High-resistivity stem	4.02	9.0×10^{-6}	862	600	435	398	---	$891{,}300\times10^{6}$	$7{,}080\times10^{6}$
HARD											
7040	G705BA	Kovar sealing	2.24	4.75×10^{-6}		702	484	450	1.480	$5{,}310\times10^{6}$	88.1×10^{6}
7052	G705FN	Weather-resist. Sealing Kovar	2.28	4.6×10^{-6}	1115	708	480	435	1.484	501×10^{6}	20×10^{6}
7070	G707DG	Low-loss tungsten sealing	2.13	3.2×10^{-6}	1100		495	455	1.469	$150{,}000\times10^{6}$	$1{,}300\times10^{6}$
7720	G702P	Nonex	2.35	3.6×10^{-6}	1110	755	525	485	1.489	653×10^{6}	15.8×10^{6}
7740	G726MX	Chemical Pyrex	2.23	3.25×10^{-6}	1220	820	565	520	1.474	141×10^{6}	4.73×10^{6}
7900		96% silica	2.18	0.8×10^{-6}	---	1500	910	820	1.458	$5{,}012\times10^{6}$	100×10^{6}

In the temperature range in which viscosities can be measured, certain standard viscosities have been selected and the temperatures reported at which the various glasses exhibit these viscosities. (See Table II and Fig. 2 for values, and below for definitions.)

Strain Point. The temperature at which the glass has a viscosity of $10^{14.5}$ poises is called the strain point. At this temperature, 90 per cent of the strain in a 1/4-inch-thick piece of glass can be removed in a four-hour period of heating.

Annealing Point. The temperature at which the glass has a viscosity of $10^{13.0}$ poises is called the annealing point. At this temperature, 90 per cent of the strain in a 1/4-inch-thick piece of glass can be removed in a 15-minute period of heating.

Softening Point. The temperature at which the glass has a viscosity of $10^{7.6}$ poises is called the softening point. Strain vanishes in a fraction of a second at this temperature.

Working Point. The temperature at which glass has a viscosity of 10^{4} poises is called the working point. At this temperature, the glass is fluid enough to be blown, pressed, or pulled.

Annealing of Glass. Because glass has high inherent viscosity, or resistance to flow, as well as poor heat conductivity, the stresses applied in the glass-forming operations, plus those arising from nonuniform cooling, tend to set up strains, which are displacements of molecular dimensions. These strains can be relieved (provided the components have equivalent expansions) as indicated above when the article is heated to the annealing range and held there for a sufficient time to allow the displacements to disappear through viscous flow. However, after the strain has been removed, it is necessary to cool the glass through the annealing range slowly and uniformly enough so that no thermal gradient exists in the glass at the time it reaches the strain-point temperature; otherwise, the removal of a thermal gradient below the strain point will reintroduce strain in the glass.

If no thermal gradient exists in annealed glass at or below the strain point, it can be cooled as rapidly as desired without danger of re-introducing permanent strain. However the glass must not be cooled so rapidly and unevenly as to break it from sudden temporary strain. After the glass has reached room temperature, this temporary strain disappears and with it the danger of breaking from this cause. Cooling rates of several degrees Centigrade per minute can safely be used in all cases of electron-tube annealing. For example, carefully made small glass tubes can be subjected to a thermal shock of 100 C temperature change in a few seconds without danger of breakage.

Because, as will be explained in the section dealing with the strength of glass, the annealed or strain-free condition is not the strongest condition for a glass article, few electron tube parts are not subjected to a rigorous annealing process.

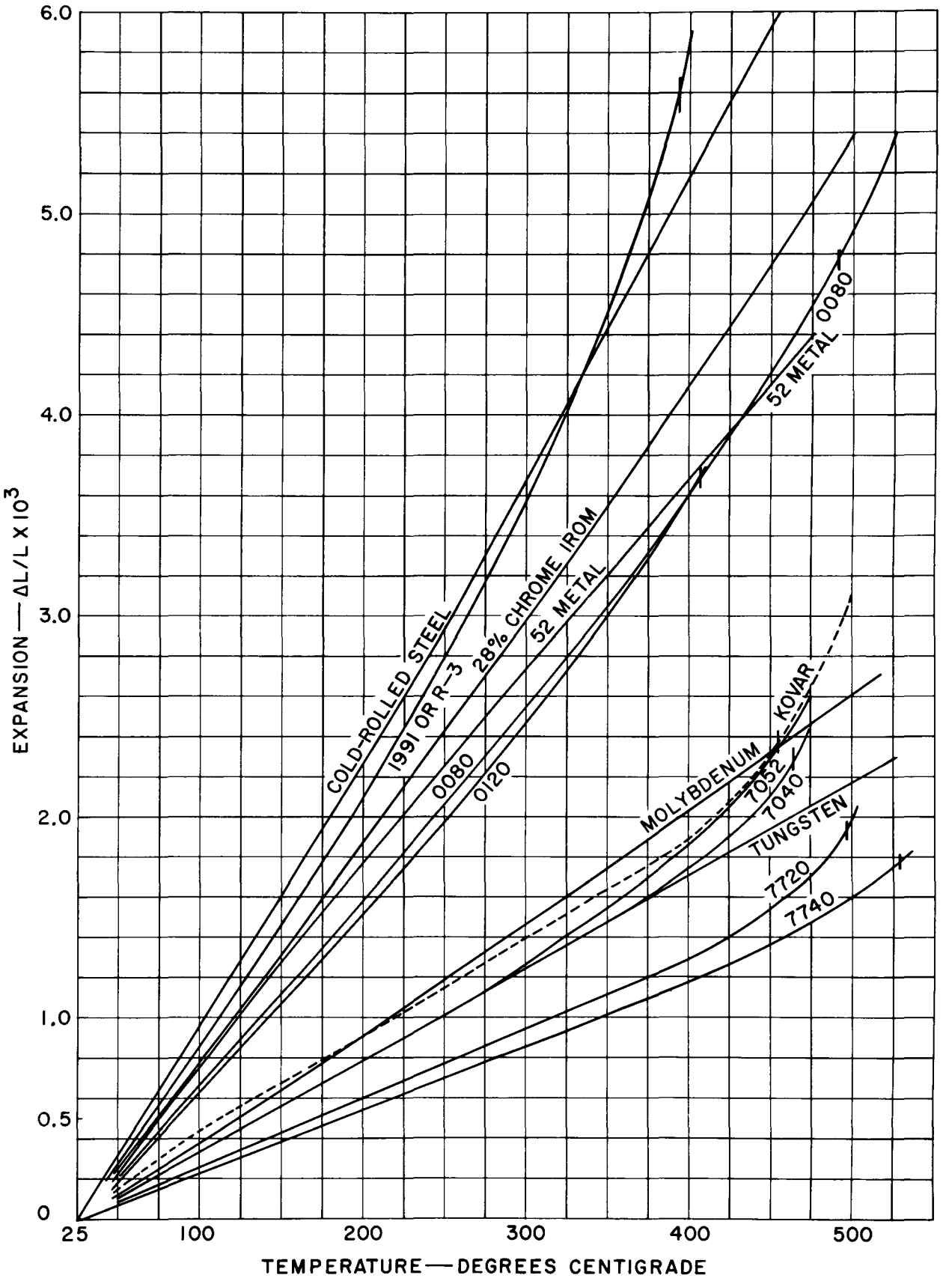


Figure 1. Thermal Expansion of Representative Glasses

Table III

Some Common Glass-to-Metal Seal Components

Glass	Group	Metals Which Match the Glass
0010, 0080, 0120, 8160, 8161	1	Platinum (not widely used because of high cost). Dumet (widely used); consists usually of 20% Cu sheath, 80% Ni-Fe alloy core (42% Ni, 58% Fe). 27% chrome iron (27% Cr, 73% Fe). Sealmet No. 4 (6% Cr, 42% Ni, 52% Fe) Allegheny-Ludlum; also has been called Driver-Harris No. 14 alloy, Sylvania No. 4 alloy, and Carpenter No. 426 alloy. 52 metal (50% Ni, 50% Fe).
1991 or G. E. R-3	2	Cold-rolled SAE 1010 steel.
7040, 7052	3	Kovar Molybdenum
7070, 7720	4	Tungsten
7740	5	Tungsten which has been beaded with 7720 or with 3320 uranium glass.
7900, Fused Quartz	6	No sealing metal.

Electrical Resistivity

Glass belongs to the class of good electrical insulators. Because glass is an undercooled liquid, the electrical conduction process is electrolytic; i. e., the current is carried by the actual movement of charged ions through the glass. The conduction current is carried mainly by sodium and oxygen ions. Ions larger than sodium or oxygen have difficulty in getting through the random network structure of glass at temperatures below the annealing point. Consequently, glasses low in sodium have high electrical resistivities, whereas glasses having a higher sodium content conduct more current at the same applied potential difference. Of glasses having different softening characteristics but the same sodium oxide content, the more refractory glass normally has the higher electrical resistivity because the current-carrying ions encounter more resistance to movement due to the greater viscosity of the more refractory glass.

The resistivity of glass varies inversely with the absolute temperature at a logarithmic rate; i. e., the logarithm of the resistivity plotted against the reciprocal of the temperature yields a reasonably straight line, as shown in the curves of Fig. 3. Resistivity values for two temperatures are also given in Table II for the glasses listed there.

The use of glasses having high electrical resistivity is important for certain electron-receiving-tube applications, not because the small electrolysis currents (normally a few hundreths of a microampere) through the glass between lead wires interfere materially with tube operation, but because a glass of low resistivity that is used as a stem can liberate gas into a tube during life and can even, in the worst conditions, rupture to allow air to enter the envelope. These possibilities arise from the fact that the electrical conduction process in glass, with its actual movement of ions, is disruptive — it changes the composition of the glass in the neighborhood of the lead wires and weakens the bond in their glass-to-metal seals. The changed composition induces a physical strain or dislocation which eventually can rupture the glass.

In the specific case of the heavy-duty rectifier stem shown in Fig. 4, sodium ions had migrated to the plate leads during the negative swing of the plate and had there reacted with the lead oxide in the glass to form black lead trees. At the same time, the oxygen from the copper oxide layer on the dumet of the plate leads migrated toward the temporarily positive filament leads. This oxygen first oxidized the leads to form brown copper peroxide, and later caused cracks along their sealed length when the strain from dissolved oxygen finally exceeded the elastic limit of the glass.

When trouble is experienced from electrolysis in electron tubes, it can sometimes be reduced if the stem is annealed before the sealing process. This additional annealing step helps reduce electrolysis because the resistivity of an annealed stem may be as much as three times that of a strained one. A badly strained stem made of 8160 glass, for example, might have no more resistance than a well-annealed stem made of the lower-resistance 0120 glass.

Ordinarily, little difficulty is experienced with electrolysis in soft glass stems when the operating temperature of the stem is 200 C and below. As an operating temperature of 250 C is approached, however, considerable evidence of electrolysis is found in any flat-press stem (made from 0120 or 8160 glass) which has passed a 500-hour life test. In fact, it is practically mandatory that the higher-resistance 8160 glass be substituted for the usual lower-resistance 0120 glass in the stems of all heavy-duty rectifier tubes, such as types 5U4G and 5R4-GY because of the high operating temper-

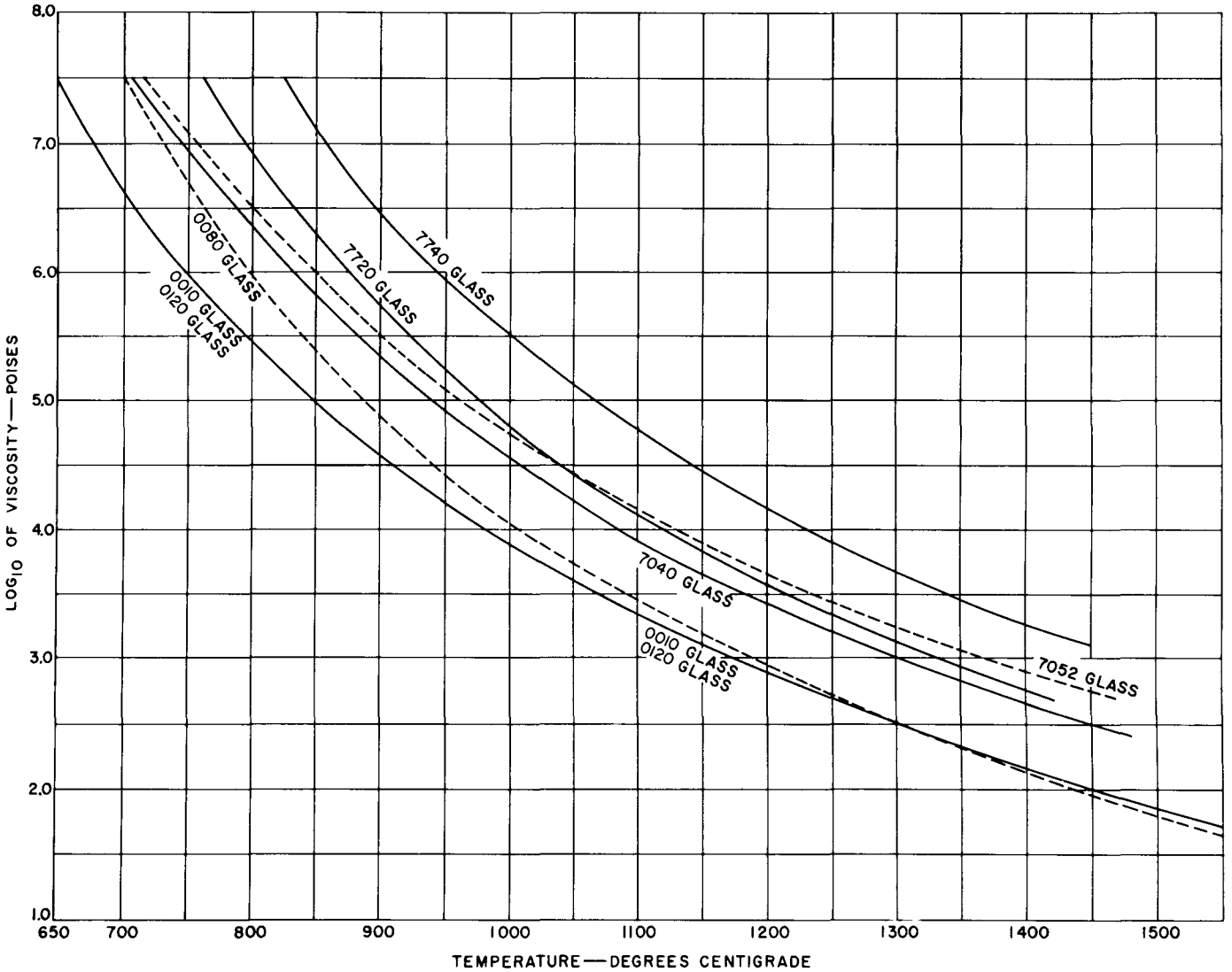


Figure 2. Viscosity vs. Temperature Curves of Representative Glasses

ature of the stems of these tubes and the high voltage across them. Morever, in many cases, even the use of 8160 glass has been found insufficient to produce tubes which will always survive a 500-hour life test with an adequate factor of safety for electrolysis failures. As a result, 8161 glass was created with a resistivity approximately 20 times that of 8160 at 250 C. Up to 1960, no receiving tube using 8161 glass has shown serious signs of stem electrolysis even after 2000 hours of life. Unfortunately, 8161 glass is more expensive than 0120 and 8160 glass, and is somewhat more difficult to work because of its shorter working range.

In many cases, stem electrolysis can be avoided if a button stem is substituted for the flat-press stem. This improvement results from the fact that button stems normally permit greater spacings between lead wires than do flat-press stems; and, in addition, button stems, being nearer the bottom or outside of the tube, run at a lower temperature than do flat-press stems.

High voltages are not always necessary to produce severe electrolysis. If stem temperatures above 250 C are experienced, severe electrolysis can occur when the potential differences between adjacent leads is no greater than 150 volts. The effect of relatively low voltages at high temperatures can be readily understood when it is noted that the amount of electrolysis current doubled for every 15 C rise in temperature in a test conducted in the laboratory on stems made from 8160 glass and heated over the range from 200 to 350 C with a constant voltage across adjacent leads.

PHYSICAL PROPERTIES OF LESSER INTEREST FOR RECEIVING TUBE MANUFACTURE

Density

The glasses used in electron tubes range in density from a low of 2.13 grams per cubic centimeter for 7070 tungsten-sealing glass, to a high of 4.02 grams per cu-

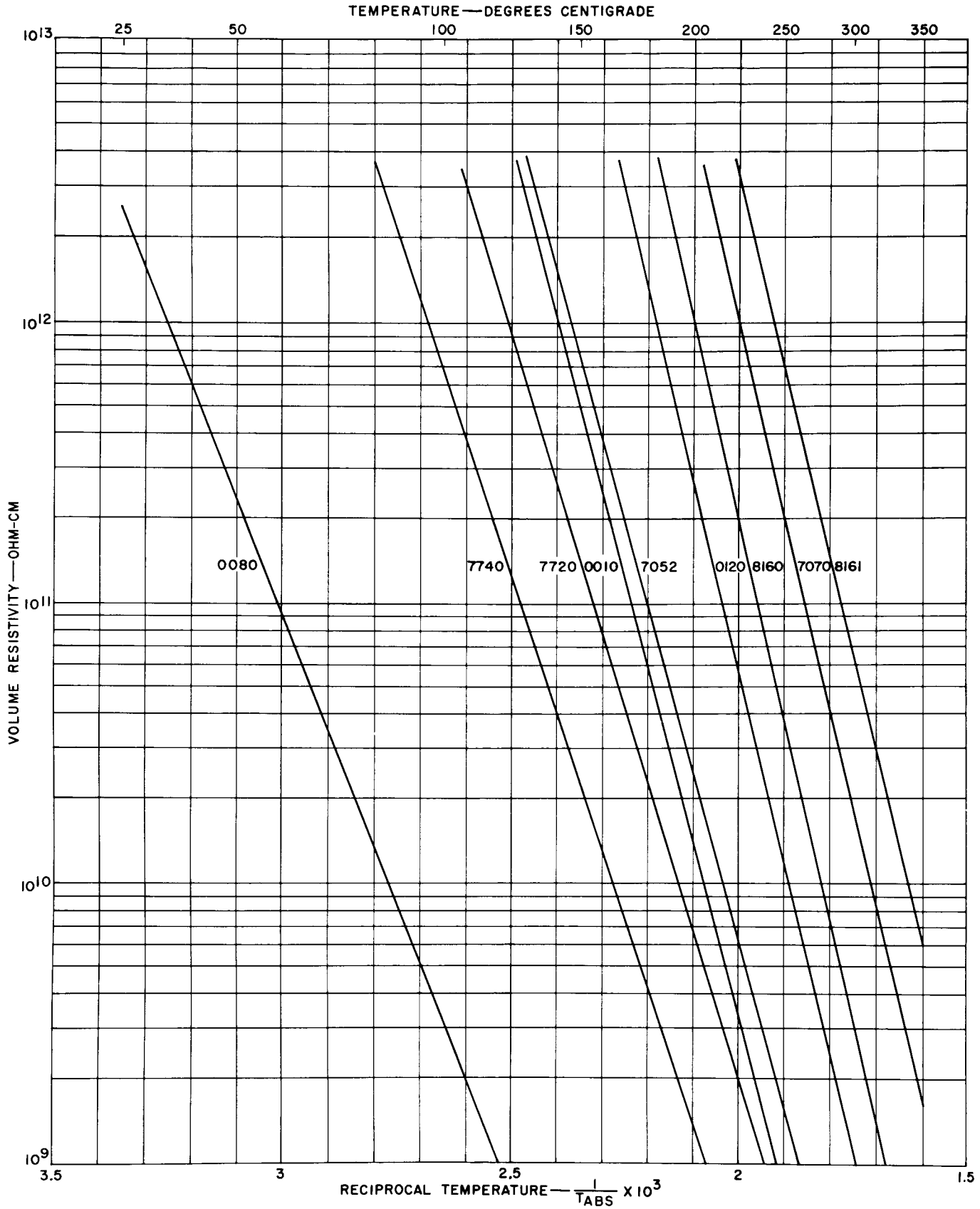


Figure 3. Resistivity Curves of Representative Glasses: Part 1, Resistivity = 10^9 to 10^{13}

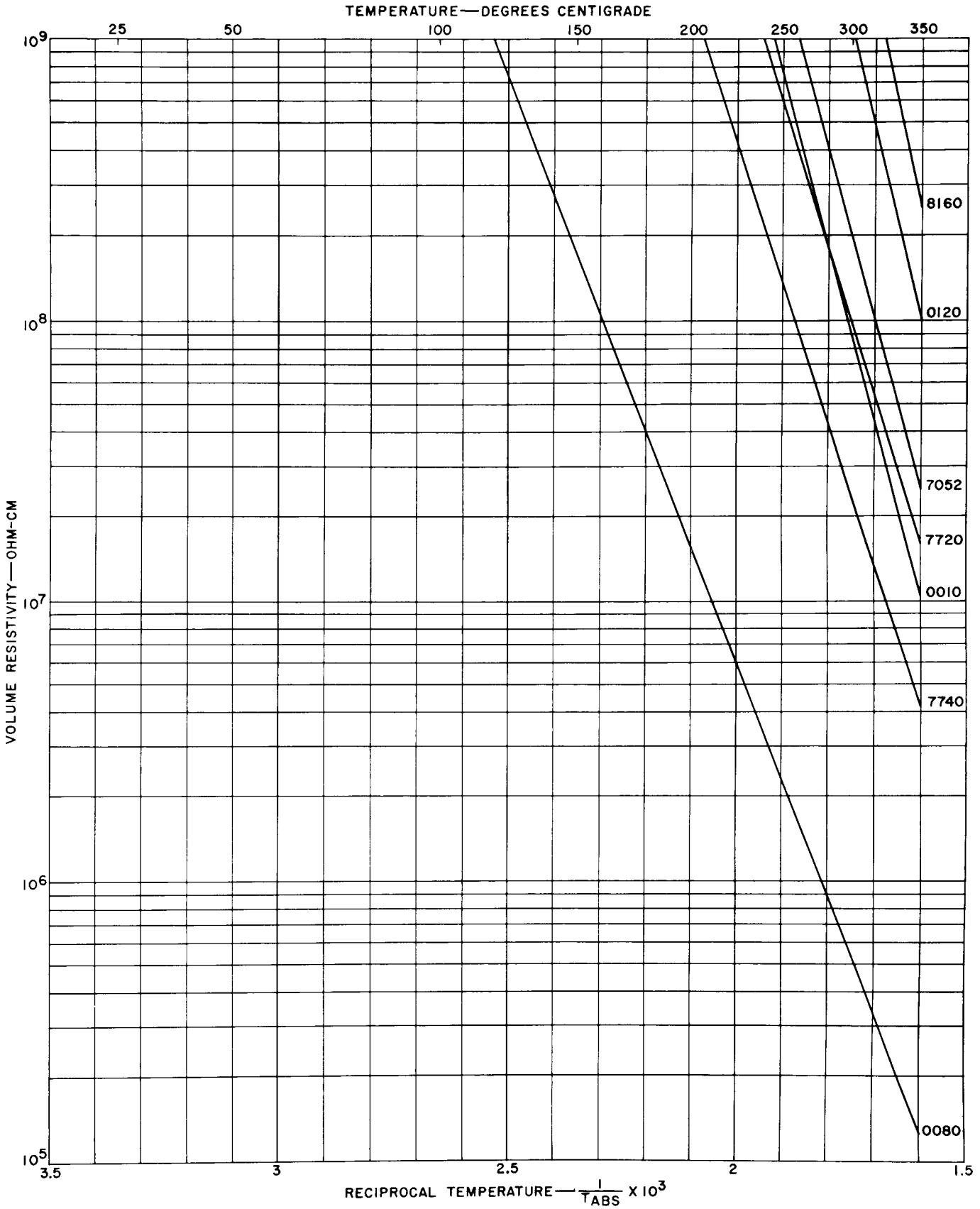


Figure 3. Resistivity Curves of Representative Glasses: Part 2, Resistivity = 10^5 to 10^9

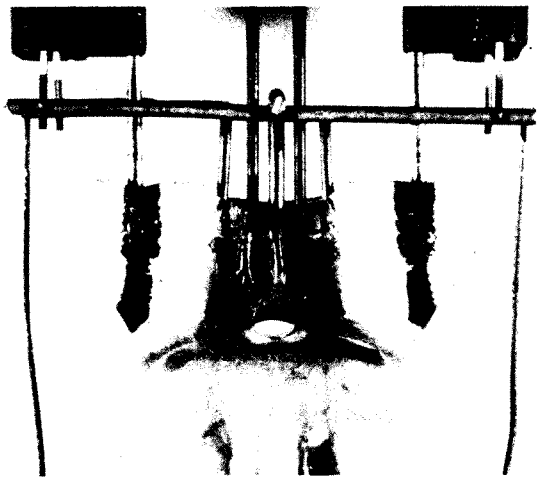


Figure 4. Electrolyzed Stem of Heavy-Duty Rectifier Tube

bic centimeter for 8161 high-resistance glass. Density is of interest mainly in that glass is purchased by the pound. Consequently, a dense glass will yield fewer stems of a given size per pound than one of a lesser density. Furthermore, this property must be used in figuring the weight to which glass beads should be cut for a given size stem. In addition, because density is easily determined with considerable accuracy, its determination provides a quick and reliable check on the identity and uniformity of glass batches and parts.

Weathering Resistance

The resistance of a glass to decomposition by moisture and atmospheric constituents is termed weathering resistance. This property varies widely for different types of glass. As a class, soft glasses are less resistant than hard glasses, although there are exceptions. Certain "solder glasses" have the poorest weathering resistance, whereas chemical Pyrex is designed to have the best resistance (next to fused silica). Laboratory tests are used to rate the glasses in different orders depending on whether acidic or basic testing solutions are employed. Of the glasses used in normal receiving-tube manufacture, 0080 lime glass bulbs are the most susceptible to damage from weathering. This weathering almost never progresses far enough to imperil the strength of the bulb but rather is harmful in that the products of decomposition affect not only the secondary-emission ratio of the glass, but also the cathode emission of the tube.

The harmful effects of weathering can be avoided by using the glass before it has time to weather, by use of a weather-resistant glass, or by the washing of a weathered glass with hydrofluoric acid.

Dielectric Constant and Power Factor

The dielectric constant of a material is defined as the ratio of the capacitance of a capacitor having that material as the dielectric to the capacitance of the same capacitor having a vacuum as the dielectric. The dielectric constants of the normal commercial glasses fall between 4 and 8 at a frequency of 60 cycles. Dielectric constant has little bearing on the use of glass in most receiving tubes but becomes important when tubes are operated at high frequencies and high voltages.

The power factor of a material is a measure of the electrical energy lost as heat in the material when it is placed in an alternating electric field. It is easily determined from the angle between the current and the charging potential. This loss varies with frequency and with the composition of the glass from as little as 0.05 per cent to as much as 6 per cent. At the frequencies and voltages normally used in electron receiving tubes, power factor is unimportant and may be safely neglected. This is not always true for power tubes. Here low-loss glasses must often be used to avoid serious overheating effects.

Table IV lists the dielectric constants and power factors of a number of representative glasses for frequencies from 60 cycles to 1 megacycle, and, in a few cases, for higher frequencies. The relatively poor power factor of lime glass is due to the high soda content (17.5 per cent Na_2O).

OPTICAL PROPERTIES — COLOR, LIGHT TRANSMISSION, LUMINESCENCE, INDEX OF REFRACTION, AND BIREFRINGENCE

Color

The color of commercial glasses other than filter glasses is normally caused by small amounts of impurities which occur naturally in the raw materials or, in some cases, which are added to the glass batch for identification. Thus, 0080 lime glass is slightly greenish when one looks through thick pieces, for example, edgewise through long tubing. The greenish cast is due to small amounts of dissolved iron oxide impurity. Correspondingly, thick pieces of 0120 lead-glass tubing are blue because of a small amount of cobalt oxide added for identification. Pyrex is yellow when viewed similarly. The color usually has no effect on any other physical property. However, when the color is black due to reduction of lead oxide to metallic lead on the surface of a lead-containing glass such as 0120, there is a reduction in the surface resistivity of the glass. The amount of this reduction, however, is not sufficient to spoil the glass for use as the insulating medium in a tube stem; tubes with blackened stems and seals operate satisfactorily.

Light Transmission, Luminescence, and Index of Refraction

The light transmission of receiving-tube envelopes and stems is of value in that it allows nondestructive visual inspection of completed tubes. The transmission of the standard clear glasses is approximately 91 per cent in the visible spectrum (wavelengths between 4000 and 7000 angstroms). The major portion of the loss in the visible spectrum results from a loss of about four per cent by reflection at each surface. However, the so-called clear glasses start to absorb energy in noticeable amounts in the ultraviolet region below 4000 angstroms. Light transmission through ordinary lead and lime bulb glasses is down nearly to zero at 3000

Table IV

Power Factors and Dielectric Constants of Some Common Glasses at 20 Degrees Centigrade

Glass	Code No.	Lab. No.	Power Factor (per cent power loss)	Dielectric Constant	Frequency
Lime Glass	0080	124HD	6.4	8.25	60 cycles
			3.0	7.65	1 kilocycle
			0.9	7.20	1 megacycle
Borosilicate	7070	707DG	0.06	3.95	60 cycles
			0.06	3.95	1 kilocycle
			0.055	4.00	1 megacycle
	7760	720G0	0.18	4.7	740 kilocycles
			0.17	4.7	1 megacycle
	7740	726MX	1.27	4.8	60 cycles
			0.77	4.7	1 kilocycle
			0.40	4.7	1 megacycle
Lead Borosilicate	7720	702P	0.38	4.9	1 kilocycle
			0.22		50 kilocycles
			0.22		732 kilocycles
Lead Glass	0010	G1	0.36	6.63	60 cycles
			0.25	6.6	1 kilocycle
			0.16	6.55	1 megacycle
			0.165		5 megacycles
	0120	G12	0.105	7.1	1 megacycle
			0.140		5 megacycles
			0.21		10 megacycles
Kovar Sealing	7040	705BA	0.22	4.76	1 kilocycle
			0.20	4.76	3 kilocycles
			0.19	4.75	10 kilocycles
			0.18	4.75	30 kilocycles
			0.11	4.75	100 kilocycles
	7052		0.2	5.1	1 megacycle
96% Silica Glass	7900		0.076	3.8	60 megacycles
			<0.05		1 megacycle
			0.064	3.8	60 cycles
R-3 Iron Sealing Glass	1991		0.086	8.13	1 megacycle

angstroms. A similar absorption occurs in the far infrared region, and complete cutoff occurs between 45,000 and 50,000 angstroms for all normal glasses. As consequences of these effects, special ultraviolet-transmitting glasses must be used for ultraviolet lamps, and electron tubes and electric lamps run hot because of the absorption of infrared, or heat, rays in their envelopes.

The luminescence of commercial glasses is also of little importance with respect to electron tubes, save as a method of identification. Under short-wave ultraviolet illumination (2537 angstroms), all high-lead glasses glow a bright blue. This effect permits ready discrimination between lead and lime bulbs because 0080 lime glass fluoroesces very weakly in comparison with lead glass.

The index of refraction, or light-bending power, of unstrained glass is important to its use in lenses and prisms, but has no bearing on its use in electron tubes. It is, however, used for identification of the glass in the laboratory. As listed in Table II, the indices of refraction for commercial glasses vary from about 1.45 to 1.80 with most of them falling in the range from 1.45 to 1.60.

Birefringence under Strain

All glasses under mechanical or thermal stress become strained, and most of them, when in this condition, exhibit birefringence; i. e., they have two or more different indices of refraction. When such a strained piece of glass is placed between crossed light-polarizing devices, interference phenomena are produced which can be interpreted both qualitatively and quantitatively to determine whether the glass may crack or rupture spontaneously. The qualitative interpretation delineates the direction and nature (tension or compression) of the strains. The quantitative interpretation gives the amount of the stresses. Therefore, the property of birefringence under strain is most useful, although it has no influence on the endurance of one glass as compared with that of another having a different birefringence constant.

Table V lists the birefringence constants for a number of representative glasses. These constants are given in millimicrons birefringence per centimeter of optical path per kilogram of load per square centimeter of cross section ($m\mu/cm/kg/cm^2$). Thus, a force of 1 kilogram applied to the opposite sides of a one-centimeter cube of 0120 glass produces a retardation, or difference between fast and slow rays, of 2.5 millimicrons of optical path. In practice, a force of 1 kilogram per square millimeter is considered to be about the safe upper limit for stress in a glass stem or seal. This amount of stress would give a reading of 250 millimicrons for the centimeter cube. Because receiving-tube stems normally work with thicknesses of about a fifth of this (0.2 centimeter), values of over 50 millimicrons begin to be dangerously high.

TABLE V

Birefringence Constants of Some Common Glasses

Code No.	Lab. No.	Birefringence Constant $m\mu/cm/kg/cm^2$
0010	G1	2.75
0080	G8	2.40
0120	G12	2.50
7070	707DG	4.35
7720	702P	3.40
7740	726MX	3.45
1991 or R-3	---	2.20
8160	814KW	2.57
7052	---	3.46

STRENGTH, CRACKS, AND STRAIN CONTROL

Because glass is an undercooled liquid, it, like other liquids, has a very high compressive strength and a considerably lower tensile strength. The tensile strength of glass is not more than one-tenth its compressive strength. This strength difference coupled with the fact that glass is a poor conductor of heat with about 1/300 the heat conductivity of metallic copper determines that the strength of glass is largely controlled by the condition of its surface. The following paragraph explains the relation of surface condition to strength.

The interdependence arises from the cooling behavior. The surface of the glass cools first from the forming, or working, temperature because it is in contact with the cool ambient air and because the heat travels slowly from the hot interior of the glass. As a result, the surface contracts and assumes its final form while the interior is still expanded and free to flow to remove stress. On continued cooling, the center portion solidifies and contracts in all directions, which has the effect of putting the surface layer under tangential compression. Because no stresses can exist at right angles to a free surface, the radial tension which forms inside the article vanishes at the surface and the surface is in pure compression in all directions. Thus, the surface of the article tends naturally to assume the strongest condition for the material.

Although this condition leaves the interior of the article in tension (the relatively weak condition), if there are no discontinuities in the interior such as air bubbles, glass stones (inclusions of crystalline matter), weld knots on leads, and the like, the stress has no focal point on which to concentrate and cannot cause a crack to start. Such discontinuities as edges, pits, or bulges are almost impossible to avoid on the outer surfaces and would serve as focal points for stress concentrations if the surface were allowed to go into tension. Hence, an article in compression on the surface and in tension on the inside will almost always resist cracking more strongly than would one in tension on the outer surface and in compression inside.

In addition, a glass article in compression on the outer surface resists cracking better than one which is neutral on the surface. The reason for this behavior is that the compressed surface cannot go into the weaker tension condition without first going through the neutral state when the article is stressed by an external force.

When a normal compressed-glass surface layer is scratched or abraded, it is weakened in proportion to the depth of the scratch or abrasion. Deep scratches which completely penetrate the compression layer may cause spontaneous failure. Parts which bear shallower scratches, when immersed in water, may spontaneously fail, possibly because of the wedging action of the water pulled into the tiny fracture by capillary attraction.

Because, in electron-tube manufacture, it is impossible to keep discontinuities (caused by the presence of lead-in wires which must go through the glass) from the interior of tube stems and seals it is not sufficient to depend solely on the natural surface compression layer of the glass to prevent cracks. In addition, it is necessary to control the radial and vertical distribution of strain around these discontinuities.

For the control of strains in the interior of glass-disc or button stems in the region of discontinuities, it must be remembered that a radial compression normally has associated with it a tangential tension, and that a radial tension is associated with a tangential compression. Further, a break or crack always travels in a direction at right angles to the direction of the principal tension in the region of the crack.

Fig. 5 depicts two button stems showing the two types of cracks normally experienced, and the strain pattern colors which usually accompany each type (standard RCA red-tint plate orientation is used).

In analyzing the cause of cracks in glass receiving-tube envelopes, it is necessary to realize that glass can crack not only from the permanent room-temperature strain which remains in the glass after cooling, but also from transient mechanically or thermally induced stresses which afterward leave no trace of their existence in the cold equilibrium strain pattern. Furthermore, there will have been no indication of strain before the treatment which caused the transient stress. A break due to a transient condition must then be analyzed on the basis of its type, point of origin, and direction of travel.

In button stems, breaks due to radial tension normally take curved paths involving the edge of the stem and one or more lead wires. Breaks due to high tangential tension usually are straight and follow a radius of the stem. The point of origin of a crack can often be determined by the presence of cracks radiating from it. Also, a study of the cracked surface under the microscope often shows a number of curved concentric lines surrounding or partially surrounding the point of origin. These lines are the locus of points on the crest of the shock wave and represent pauses of the shock wave during the cracking process. Fig. 6 shows such a break pattern. If it is remembered that glass is an undercooled liquid, it is easy to understand the analogy of these lines to the concentric wave crests which surround a pebble dropped on the surface of a pool of still water. The origin of the crack always lies behind the hollow or concave side of the lines and the crack always progresses in the direction of the full or convex side of the lines.

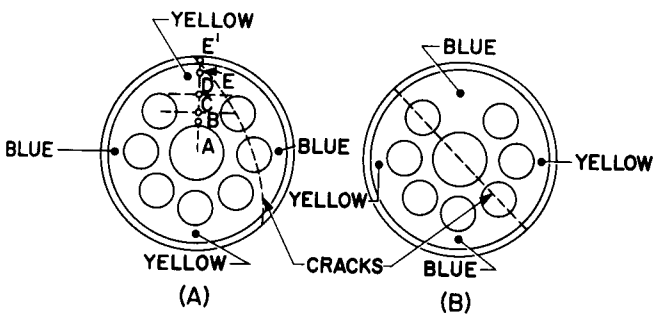


Figure 5. Strain Patterns and Cracks in Button Stems (colors are those obtained with first-order red-tint plates, standard RCA orientation); (A) Radial tension pattern (radial tension type crack. Letters show standard polarimeter measuring— points per Standardizing Notice 34-37-12A); (B) Radial compression pattern (Tangential tension type crack, often called a radial compression crack)

THERMAL AND MECHANICAL TESTS

Because the strength of glass, as previously explained, resides in the surface and is materially affected by anything which affects the shape, contour, discontinuities, flaws, and strain condition of the surface, it is a complex and highly variable property. Because it would be very difficult to measure the effect of all these numerous variables separately, a test which blankets as many as possible has been devised for miniature-type tubes. This test measures the ability of the tube envelope to endure a double thermal shock while the base pins of the tube are spread as they would be in a tight tube socket having a large pin-circle diameter. In brief, the tubes are placed over a truncated metal cone which spreads the leads five degrees. The tubes are then plunged into boiling water for a prescribed period of time, and then plunged into ice water immediately afterward. Various combinations of this mechanical and thermal-shock test have been tried and found to give practical control of the quality of tubes made and shipped.

Because of the deterioration of glass-surface strengths with time, the strength test applied at the time of manufacture is more stringent than the tube should be expected to withstand weeks or months later. Also, because the strength of glass is such a complex property and varies widely from specimen to specimen, samples must be tested on a statistical basis according to logical sampling plans.

The foregoing explanation is not meant to imply that individual measurements of strength cannot be made and used advantageously. In fact, the laboratory has a number of such methods for the measurement of bulb strength, lead pull-out strength, tip strength, and the

Figure 6. Cracked Section of Flat-Press Stem

like. However, because of the need for very careful control and because of the wide variability encountered, such tests are not normally used in factory control work.

CHEMICAL TREATMENT OF GLASS SURFACES

In recent years a number of chemical treatments have been developed to protect the surface of glass and to improve its durability.

Ammonium stearate is currently being used by glass manufacturers in the annealing lehr to form a lubricating film on the surface of the glass which reduces abrasion and breakage during shipment. Because ammonium stearate is completely volatile in the 200 to 300 C temperature range, it evaporates completely when the bulbs are sealed.

Sulfur dioxide has been used by the bulb manufacturer during the bulb annealing process to toughen and lubricate the surface of the bulbs. Unlike ammonium stearate, sulfur dioxide reacts with the glass, and some sulfur remains after the tube-sealing operation. The presence of sulfur has been blamed for lowered cathode emission and sulfur dioxide is not now used.

Weathered glass bulbs can be washed with 1 or 2 per cent hydrofluoric acid to give a fresh glass surface. The mechanism is the formation of silicon tetrafluoride which is evolved as a gas to leave a new surface after other decomposition products are rinsed away.

Silicone treatment of a glass surface lubricates the surface so that it is not abraded or damaged by contact with other glass surfaces. In addition, the silicone coats the glass with a water-resistant coating which greatly inhibits the normal deterioration of glass strength with time. The disadvantage of silicone coatings for finished tubes is that the coatings must be cured at temperatures of about 300 C to be effective. This temperature is more than the average finished tube can stand without becoming gassy which impairs its electrical performance.

Hydrogen firing of lead bulbs forms a coating of reduced lead on the bulb and renders the glass very resistant to decomposition under high-voltage bombardment. This process also toughens the glass to mechanical breakage. As a result, the process is currently used on rectifier tubes whose bulbs are normally bombarded by electrons accelerated by voltages of 20,000 volts or more.

SOME GENERAL CONSIDERATIONS CONCERNING GLASS-STEM, BULB, AND SEAL DESIGN AND USE

With soft-glass stems using dumet as lead-wire material, it has been found that the lowest losses due to cracks around the lead wires occur if the lead diameters are kept at or below 0.020 inch. Larger lead diameters, up to 0.040 inch, can be used if care is taken in handling the stems. However, all dumet-to-soft-glass seals are always in a state of strain because of the thermal expansion mismatch caused by the two different expansions of dumet: radial and axial. The magnitude of this strain increases until it becomes so great above 0.040-inch wire diameter that most of the seals will break.

Glass-to-metal seals should never be designed to be made in a reducing atmosphere, nor should finished stems be annealed in a strongly reducing atmosphere for more than a minute or two. Because an oxide bond between the glass and metal is necessary for true chemical adherence, seals which are made in reducing atmospheres will not normally be vacuum-tight.

Distances between adjacent lead wires in a stem should be at least three or four times greater than the wire diameter. Closer spacing between leads usually results in excessive cracks in the close-spaced region.

Lead wires should always be located in a stem to provide the maximum symmetry. Lack of symmetry produces unbalanced strain which tends to cause cracks.

Lead wires which go through bulb walls or domes should always be beaded with a glass which matches the thermal expansion of the bulb. The beading can be done either before sealing or coincident with the sealing operation.

Different glasses, having diameters greater than one inch should not ordinarily be sealed together. Graded seals of tubing above this size are increasingly subject to cracking unless the glasses match closely from room temperature to the setting point. (Graded seals are multi-component seals comprising glasses of different expansions sealed in order of increasing or decreasing thermal expansion. They are used to join two glasses of widely different expansions.)

In addition, it is desirable to avoid sealing different glasses together if their softening points are radically different. Even though the straight-line portions of the expansion curves may match, the setting points and the expansions to the setting points will not match (see Fig. 1). Furthermore, the glasses will not diffuse into each other to make a strong seal.

Glass tubes should not be mounted in contact with hot external metal surfaces because the glass will be highly strained at the point of contact due to the poor heat conductivity of glass. Glass tubes having mount structures which permit bombardment of inner bulb walls with high-voltage electrons during tube operation should be mounted with their bulb walls at least a half inch away from the nearest grounded metal surface (filament supply lines, metal shields, or cages). If this precaution is not taken, the high-voltage charges on the glass leak through to the ground and heat the glass. More current, therefore, flows through the lowered resistance of the hot glass, and the glass may finally be perforated leaving a heat-polished hole. All high-voltage and heavy-duty rectifier tubes are prone to this type of failure.

In tubes which operate at relatively high temperatures, such as heavy-duty rectifier tubes, care should be taken to place the getters either at the top or bottom of the mount structure so that the gettering material will be deposited either on the lower shoulder of the glass bulb or in the dome of the bulb. This procedure leaves the sidewalls opposite the hot plate structure

free to transmit some of the heat to the outside. If these sidewalls are covered with mirror-like getter deposits, they will reflect the heat back into the mount structure and raise the operating temperature of the tube.

Getters should never be placed so that the getter film is deposited on the sides of the stem press. If the getter film does not tend to cause leakage between the electrodes in the stem press, it may still act as an electrode which collects back emission from the plates and which promotes electrolysis between getter film and lead wires.

REFERENCES

Books

Weyl, W. A., Coloured Glasses, The Society of Glass Technology, Sheffield, England, 1951

Littleton, J. T. and G. W. Morey, Electrical Properties of Glass, J. Wiley & Sons, New York, 1933

Shand, E. B., Glass Engineering Handbook, McGraw-Hill, New York, 1958

Day, R. K., Glass Research Methods, Industrial Publications, Chicago, 1953

Phillips, C. J., Glass: The Miracle Maker, Pittman Publishing Corp., New York and Chicago, 1941

Partridge, J. H., Glass-to-Metal Seals, Society of Glass Technology, Sheffield, England, 1949

Scholes, S. R., Handbook of the Glass Industry, Ogden Watney Publishers, New York, 1941

Tooley, F. V., Handbook of Glass Manufacture, Ogden Publishing Co., New York, 1953

Stevels, J. M., Progress in the Theory of the Physical Properties of Glass, Elsevier Publishing Co., New York, 1948

Morey, G. W., Properties of Glass, Reinhold Publishing Corp., New York, 1938

Barr, W. E. and V. J. Anhorn, Scientific and Industrial Glass Blowing, Instrument Publishing Co., Pittsburg, 1949

Hodkin, F. W. and A. Cousen, Textbook of Glass Technology, Constable & Co., Ltd., London, 1925

Scholes, S. R., Modern Glass Practice, Industrial Publications, Chicago, 1946

Periodicals

Glass Industry
Bulletin of the American Ceramic Society
Journal of the American Ceramic Society
Journal of the Society of Glass Technology

Glass and Its Properties and Seals

J. C. Turnbull and G. E. Eiwen

Lancaster

Glass is an essential material used in making the envelopes of many electron tubes. It serves as a part of the mechanical structure of the envelope to provide an enclosure which not only can be made vacuum tight but which has usefulness structurally and as an electrical insulator; in addition, its optical characteristics permit light or radiation to enter or leave the tube. In some cases, glass is used for internal elements in tube construction.

The proper functioning of different tube types requires the use of many glasses of widely different physical and chemical properties. Glasses are almost invariably incorporated into tube structures by means of glass-to-metal or glass-to-glass seals and, as a result, the effects of sealing methods on the required glass properties must be considered. In addition, the specialized high-temperature processing used during the assembly of some electron tubes introduces special requirements on the high-temperature properties of glasses. As the electron-tube industry has grown, there have been increasing needs for many new and specialized glasses. Since 1945, over 20 substantially different glass compositions have been put in use by the electron-tube industry. Along with this specialization in composition and properties, there have been demands for higher-quality glass. Major developments in the melting of glasses have enabled the glass industry to meet these demands for improved quality at lower cost. Finally, the tube industry has required glass in a wide variety of shapes and sizes with improved dimensional tolerances. These needs have been met again and again by the glass industry by the use of new and improved techniques for the fabrication and forming of glass parts.

The glass color kinescope is an example of recent technological progress in the glass and electron-tube industries. This tube was developed using six different glass compositions, four of which were new glasses. Among these were a non-browning faceplate glass, an X-ray-absorbing funnel glass, and a devitrifying soft-solder glass for low-temperature sealing of the panel and funnel. The 21-inch diameter glass panel of the color tube weighs 22 pounds; it is accurately pressed and has an inner surface with a maximum deviation from a prescribed surface of less than 0.023 inch. The glass funnel, also 21 inches in diameter, is made by centrifugal casting (spinning), a process developed for making large articles of uniform wall thickness and minimum weight. The sealing technique, which uses a solder glass, is a new process for making seals at 445 C— a temperature far below those considered necessary, until recently, for forming vacuum-tight glass-to-glass seals. The design, manufacture, and application of this tube depended heavily upon improved technology developed by and with the glass manufacturer.

For the future, continued progress in the applications of glass to electron tubes is anticipated. More new glasses will be required which have a wider range of physical properties and greatly improved chemical, electrical, and thermal stability. Glasses which show electronic rather than ionic conduction will be desired for some applications. Current developments, which doubtless will be exploited, will include the devitrifying glass compositions and the refractory high-alumina glasses which tend to bridge the gap in properties between crystalline ceramic materials and glasses. Glasses or glassy materials with even more novel characteristics may become available for future applications.

COMPOSITION AND PROPERTIES

Glass is an amorphous (non-crystalline) material that is rigid at ordinary temperatures and fluid at elevated temperatures. Glass melts are characterized by exceptionally high viscosity at the temperature where crystallization occurs, so that the speed of crystallization is very slow. Glass can thus be considered as a liquid which has cooled through a freezing range without crystallizing, and whose viscosity has increased on further cooling until at room temperature it has the physical properties of a rigid solid.

Most commercial glasses are silicate glasses and owe their glassy characteristics mainly to silicon dioxide. These glasses normally contain 50 per cent or more silica by weight. Other oxides are present in various proportions and can be adjusted to provide a wide variety of physical and chemical properties. The most important of these materials are the alkali-metal oxides (Na_2O, K_2O, and Li_2O), the alkaline-earth oxides (CaO, MgO, and BaO), and other oxides such as PbO, B_2O_3, and Al_2O_3. Glasses normally contain small amounts of other elements present either as additives deliberately added or as impurities. Arsenic or antimony oxides are fining agents which may be added in amounts of less than 1 per cent to reduce the size and number of bubbles present in the glass

melt. Fluorine and chlorine are similar additives which may be used to improve glass quality. The oxides of iron, titanium, and similar metals are common impurities which tend to color the glass. Cobalt and nickel oxides are coloring additives used to impart a gray color to the faceplate of TV picture tubes for improved picture contrast. Small quantities (0.5 to 1 per cent) of cerium oxide may be added to some glasses to improve their resistance to browning under X-ray exposure.

Compositions of Commercial Glasses

Approximate compositions of some commercial glasses used in electron tubes are shown in Table I. The development of a commercial glass involves the manipulation of many possible combinations of oxide ingredients, from which a composition is found with suitable physical and chemical properties. In the manufacture of glass, its properties are maintained at their desired values by minor adjustments of the batch ingredients from which the melt is made. In the purchase of commercial glasses, the properties desired rather than the exact composition of the glass are usually specified. The compositions given in Table I are the results of analyses on individual samples and, therefore, are presented only as approximate values.

Properties of Commercial Glasses

The usual physical properties given for glasses relate to viscosity at elevated temperatures, thermal expansion, density, electrical resistivity, and stress the optical coefficient. Table II gives some of these properties according to the definitions which are briefly described below. (For a more extensive discussion of the definition and measurement of these physical properties of glass, see reference 1.)

Glasses used in electron tubes are frequently classified as "hard" or "soft", depending on their working temperature. The terminology is relative rather than precise but is in common use. The higher the working temperature of a glass, the "harder" it is considered;

Table I

Approximate Percentage Compositions of Commercial Glasses

Glass Number	Silicon Dioxide	Aluminum Oxide	Sodium Oxide	Potassium Oxide	Calcium Oxide	Magnesium Oxide	Boron Oxide	Lead Oxide	Ferric Oxide	Barium Oxide	Fluorine	Arsenic Oxide	Antimony Oxide	Lithium Oxide	Cerium Oxide	Zinc Oxide	Cobalt Oxide	Nickel Oxide
0120	54.5	1.8	4.2	8.0	0.2	0.2	-	31.2	-	-	-	-	-	-	-	-	-	-
0080	73.6	1.0	16.0	0.6	5.2	3.6	-	-	-	-	-	-	-	-	-	-	-	-
0083	73.7	0.9	16.8	-	4.6	3.5	-	-	0.02	-	-	0.01	0.26	-	-	-	-	-
0010	63.0	0.6	7.6	6.0	0.3	0.2	0.2	21	-	-	-	-	-	-	-	-	-	-
8160	56.4	1.8	2.9	9.9	0.5	0.1	-	22.5	0.07	5.5	0.1	0.3	-	0.14	-	-	-	-
9010	67.3	5.5	8.1	6.3	-	-	-	-	-	11.9	0.2	-	0.4	-		0.3	-	*
K-52	66.6	3.5	6.6	7.0	1.6	0.1	-	4.8	-	8.1	0.7	-	-	0.7	-	-	-	*
K-30	67.0	0.9	4.8	10.7	-	-	-	11.8	-	4.4	-	-	-	-	-	-	-	*
0129	60.2	1.9	5.3	12.7	2.7	2.2	1.0	9.5	0.03	3.4	-	0.4	0.7	-	-	-	-	-
TM-3	64.5	4.2	8.3	9.6	2.0	0.5	-	-	-	10.2	0.4	-	-	-	-	-	-	-
TM-5	68.1	3.8	6.4	7.6	0.1	-	-	-	-	12.6	1.2	-	-	0.4	-	-	-	*
3720	69.1	3.1	16.5	0.4	7.4	2.7	-	-	0.1	-	-	0.05	-	-	-	-	0.003	0.02
5533	66.9	1.7	8.4	6.5	3.5	-	-	-	0.06	12.0	-	-	-	-	0.3	-	-	*
R-3	47.2	2.1	3.9	14.5	-	-	-	31.7	0.08	-	-	-	-	0.6	-	-	-	-
LL117	52.0	3.5	6.2	12.8	0.8	0.3	-	12.5	-	11	-	-	0.6	0.4	-	-	-	-
7052	61.0	7.7	2.2	2.1	0.2	0.4	22.3	-	-	2.7	-	-	-	-	-	-	-	-
7720	73.0	1.7	4.4	-	-	-	14.0	5.7	-	-	-	-	-	-	-	-	-	-
7750	67.3	1.7	4.6	1.0	-	0.2	24.6	-	-	-	-	-	-	-	-	-	-	-
7740	80.5	2.2	3.8	0.4	-	-	12.9	-	-	-	-	-	-	1.2	-	-	-	-
7070	70.0	1.1	0.5	-	-	-	28.0	-	-	-	-	-	-	1.2	-	-	-	-
Vycor	96	1	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-
Fused Silica	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

*Present

Table II

Physical Properties Of Commercial Glasses

Glass Number	Softness Temp. deg C	Anneal. Point deg C	Strain Point deg C	Exp. Coeff. per deg C	Density gm/cc	Log Resistivity $\log_{10}$ ohm-cm	
						250 C	350 C
0010	626	428	397	91.0×10^{-7}	2.85	8.9	7.0
0080	688	508	478	95.5	2.485	6.4	5.1
0120	630	435	400	89.0	3.05	10.1	8.0
0128	680	486	443	95.0	2.84	9.8	7.9
0129	676	488	450	98.5	2.71	9.3	7.4
1720	915	715	670	42	2.53	11.4	9.5
1990	496	360	330	127	3.47	10.1	7.7
3320	780	540	495	40	2.29	8.6	7.1
3459	681	487	454	90	2.69	-	-
3590	713	527	499	91	2.51	-	-
3720	706	524	491	96	2.52	-	-
4442	652	485	456	117	2.54	-	-
5533	694	501	471	94.5	2.64	-	-
7040	702	484	450	47.5	2.24	9.6	7.8
7050	703	500	460	46.0	2.25	8.8	7.2
7052	708	475	438	46.0	2.28	9.2	7.4
7055	718	508	474	51.5	2.29	10.2	8.4
7056	718	509	475	51.0	2.28	10.5	8.5
7070	746	490	455	32	2.13	11.2	9.1
7720	755	518	484	36	2.35	8.8	7.2
7740	820	555	515	32.5	2.23	8.1	6.6
7750	704	467	431	40.5	2.19	9.5	7.7
7760	780	515	475	34	2.23	9.4	7.7
7780	763	498	464	37	2.18	10.0	8.2
8160	627	435	395	91	2.98	10.6	8.4
8161	601	434	403	90	-	11.9	9.8
9010	650	442	411	88.5	2.59	8.9	7.0
9019	675	487	455	99	2.59	-	-
K-30	660	442	410	90	2.67	9.3	7.3
K-52	655	447	415	89	2.64	9.0	7.0
TM 5	655	450	418	90	2.63	8.9	6.9
LL2022	650	449	408	89	2.61	9.6	-

conversely, the lower its working temperature, the "softer" it is. The terminology bears little relation to the actual hardness or abrasion resistance of the glasses in question. Generally, the soft glasses are used in receiving-type tubes with low-cost lead-in or sealing materials. Hard glasses are more generally used in specialty tube types and require relatively expensive lead-in or sealing materials.

Another class of glasses has recently come into limited use and is known as "solder glasses." These glasses have lower working temperatures than soft glasses and are used to seal other materials together at relatively low temperatures (400 to 700 C).

Softening Point is the temperature at which glass has a viscosity of $10^{7.6}$ poises. In this temperature range, glass will deform noticeably under its own weight.

Annealing Point is the temperature at which glass has

a viscosity of $10^{13.4}$ poises. At the annealing point, the internal stress caused by rapid cooling from working or forming temperatures may be substantially removed in 15 minutes.

Strain Point is the temperature at which glass has a viscosity of $10^{14.6}$ poises. At the strain point, the internal stress may be removed in four hours. It may also be considered the highest working temperature of annealed glass.

Thermal Expansion Coefficient is the average increase in length per unit length per degree centigrade change in temperature over the range of 0 to 300 C. Because the expansion coefficient is affected slightly by annealing, the values given are for annealed glass.

Density is the weight per unit volume, expressed in grams per cubic centimeter.

Volume Resistivity can be described as the electrical resistance of a sample in the form of a 1-centimeter glass cube. Values are listed as the $\log_{10}$ of volume resistivity in ohm-centimeters.

Stress Optical Coefficient is the number of millimicrons of retardation (relative wave length shift for light rays polarized in the two directions of principal stress) produced in glass one centimeter thick by an applied stress of one kilogram per square centimeter. The coefficient is used to compute stresses in glass after the retardation caused by the stress is measured. It is necessary to take the stress optical coefficient into consideration when viewing glasses of different compositions in a polariscope because glasses having higher coefficients will appear to have higher stresses.

Description and Application of Glasses

A brief description of the principal glasses used in the electron-tube industry follows: their characteristics and some typical applications are discussed under five group headings. They are:

1. Dumet-sealing glasses
2. Chromium-iron alloy sealing glasses
3. Iron-sealing glasses
4. Kovar-sealing glasses
5. Tungsten-sealing glasses

Suitable glasses for a new application usually can be selected from commercially available types. Because good quality glass parts generally cannot be made on a small scale, a considerable volume of business is required to justify special production of new compositions.

Dumet-Sealing Glasses. The following glasses are used for sealing to dumet leads, or to metals of similar expansion characteristics, such as Sylvania No. 4 alloy, type 446 and 430 chromium-iron alloys, and 52 metal.

0120 lead glass is almost universally used for stems because it represents an optimum combination of the properties of high electrical resistivity, good workability for stem-making, and an extended low-temperature annealing range highly suitable for stem sealing and flame annealing of stem-bulb seals. It is used as neck tubing on many types of cathode-ray tubes. Some receiving tubes (e.g., deflection tubes and rectifier types) have been made using 0120 bulbs to eliminate electrolysis effects. Early television bulbs (10- and 12-inch types) were made with 0120 pressed glass parts; 0120 is fluoride-free and shows fairly good resistance to weathering on storage.

Corning 8160 lead glass is similar to 0120 but is modified to have higher electrical resistivity. It is used as a replacement for 0120 (at higher cost) in stems which operate at high temperatures and show electrolysis effects. It is available only in the form of tubing.

General Electric HR glass is similar in composition and properties to Corning 8160 glass.

Corning 8161 lead glass may be used in place of 8160 where exceptionally high electrical resistivity is required, but otherwise is less practical from the standpoint of cost, properties, and availability.

0010 lead glass has electrical resistivity lower than 0120 and may be used in place of 0120 where lower resistivity can be tolerated, e.g., in the exhaust tubulation of small stems; lead content and cost are lower than for 0120.

0080 glass is a soda-lime-silica composition manufactured on a large scale in the form of bulbs and tubing; it is used mainly for the bulbs of small tubes. Because the electrical resistivity of this glass is about the lowest of commercial glasses, it has poor electrical insulating properties. Chemical durability is low; the glass may develop an objectionably high surface-leakage conductance under high humidity conditions. This glass is similar in composition to that used for windows and bottles. The General Electric Co. holds its 0080 glass at a somewhat higher thermal expansion value than Corning does.

Corning 0083 glass has a composition similar to Corning 0080 glass but is melted for optical use. It is used for making the thin glass target (5 microns thick) for the image-orthicon camera tube.

Kimble R-6 glass is a soda-lime-silica composition similar to 0080 but with somewhat higher annealing point; it is used for the bulbs of small phototube types.

Forco Window Glass (FR-60) is a soda-lime-silica composition manufactured as drawn sheet glass. Small round sections of it are sealed to 0120 or 0080 tubing to make a flat-ended bulb for phototubes and similar types. It was the first glass used for the faceplates of the 16-inch round metal kinescope; for this tube type, it was sealed to a 446 alloy shell.

Corning 9010 glass is lead-free neutral-tint glass used for the panel and funnel sections of black-and-white television picture tubes. It was developed to replace 0120 in the manufacture of these large bulbs and has similar properties but with somewhat lower electrical resistivity. It contains approximately 1 per cent of fluoride, which improves melting quality but, in addi-

tion, contributes to a lower annealing range.

Kimble K-30 lead-barium glass was formerly used by this company for picture-tube bulbs. It has properties similar to 0120 with lower lead content and lower cost. K-30 is a fluoride-free glass.

Chromium-Iron Alloy Sealing Glasses. Glasses which are used with the chromium-iron alloys (such as 446 and 430) generally have slightly higher thermal expansion than the dumet group described earlier. However, because these two groups overlap in sealing properties, there is usually compatibility in sealing the individual glasses together.

Corning 9019 glass is a neutral-tint lead-free glass similar to 9010. It has higher thermal expansion, seals strain-free to 430 alloy, and has a higher annealing point. It contains cerium oxide and is a non-browning composition. In addition, this glass contains about 0.5 per cent fluoride. Glass 9019 is used for the panel of the 21CYP22 color kinescope.

Corning 0129 is a lead glass whose thermal expansion and annealing point match those of 9019 glass. It contains about 0.5 per cent fluoride. The 0129 is used for the funnel of the color kinescope where a lead glass is desired for X-ray absorption; in this application it is sealed to the 9019 panel with a devitrifying type of solder glass (Corning 7572 frit).

Corning 0128 lead glass is similar to 0120 but is modified to have thermal expansion and annealing point intermediate between those of 0129 and 0120. It is used for neck tubing on the color kinescope.

Pittsburgh Plate 3720 is a neutral-tint soda-lime-silica glass made in the form of drawn-sheet window glass. It was used for the faceplates of the metal kinescopes; for this application, its thermal expansion was adjusted to seal with moderate compression to 430 alloy. It is now used for the safety glass in many television receivers.

Pittsburgh Plate 5533 is a neutral-tint barium glass made in the form of drawn-sheet window glass. Glass 5533 contains cerium and is a non-browning composition. It is used for the faceplate of the 21AXP22 metal color kinescope and, like 3720 glass, its thermal expansion is adjusted for sealing with moderate compression to 430 alloy. This glass is fluoride-free.

Lancaster Lens G12-X1 is a modification of 0120 having higher thermal expansion for sealing to 430 alloy. A 6-inch-diameter pressed funnel of this glass is used for sealing to the metal funnel of the 21AXP22 color kinescope.

Iron-Sealing Glasses. Because iron or cold-rolled steel have high thermal expansion, they can be used for glass-to-metal seals only with special iron-sealing glasses. These glasses have thermal expansions outside the range of those normally attained in commercial silicate compositions; attainment of sufficiently high thermal expansion for this application requires that the chemical durability of the iron-sealing glasses be sacrificed to some degree.

Corning 1991 lead glass is a relatively soft (low annealing-point) glass of moderately high resistivity which seals with low strain to cold-rolled steel. It is available in the form of tubing.

General Electric R-3 glass is similar in composition and properties to 1991 glass.

Pittsburgh Plate 4442 glass is a neutral-tint soda-lime-silica composition made in the form of drawn-sheet glass; it has low electrical resistivity. It was used for the faceplate of the metal kinescope types which had cold-rolled steel shells; its thermal expansion is adjusted to seal with moderate compression to cold-rolled steel.

G-264 lead glass is an iron-sealing composition of moderately high electrical resistivity. It was used for the pressed glass funnel which was sealed to a cold-rolled steel shell on some metal kinescope types. It was also supplied in the form of tubing for the neck and stem components of these tube types. This glass is also known as Lancaster Lens LL-117CL or General Electric IS115 glass.

Kovar-Sealing Glasses. The constituents of kovar (nickel, cobalt, and iron) are adjusted to provide a glass-sealing alloy with low thermal expansion. The following glasses were developed to fit this alloy.

Corning 7040 is the oldest kovar-sealing glass. It has moderately high electrical resistivity and is fluoride-free. This glass is somewhat unstable on heat treatment, has poor chemical durability, and tends to be attacked by the cleaning and plating solutions commonly used on tubes and tube assemblies with kovar seals.

Corning 7052 was developed for applications requiring improved stability and chemical durability. It is the most widely used kovar-sealing glass and has moderately high electrical resistivity; it contains significant amounts of chloride and fluoride.

Corning 7056 is a kovar-sealing glass melted for optical use and is used for small faceplates requiring high optical quality. It is an improved glass, is stable, has high electrical resistivity, and is free of fluoride and chloride. Commercially melted 7056 tubing is being investigated for possible application to electron tubes.

Kimble K-650 is a kovar-sealing glass having low electrical resistivity and resistance to browning by X-rays. It has a limited application for electron tubes.

Bausch and Lomb BSC-51 is an ultraviolet-transmitting glass used for small bulbs. Its thermal expansion is lower than that of kovar sealing glasses. This glass will seal to a 7052 stem and is available only in the form of tubing.

Corning 9741 is an ultraviolet-transmitting glass used for small bulbs. Its thermal expansion is lower than that of kovar sealing glasses. This glass will seal to a 7052 stem and is available only in the form of tubing.

Tungsten-Sealing Glasses. Tungsten has the lowest thermal expansion of the metals used for glass sealing. The borosilicate glasses of lower thermal expansion to be described are generally used in applications involving tungsten seals.

Corning 7720 (Nonex) is a lead borosilicate glass which has moderately high electrical resistivity. It is available in the form of bulbs, tubing, and pressed parts. This glass is the one most generally used for stems with tungsten leads.

Corning 7750 is a tungsten-sealing glass having moderately high electrical resistivity and somewhat higher thermal expansion than 7720. It may be used as an intermediate glass in sealing 7720 glass to 7052 glass. It is used for stems having large-diameter tungsten leads.

Corning 7740 is the chemical Pyrex glass used for laboratory glassware and thus is generally available. It may be sealed with some difficulty to 7720 glass and is not usually sealed directly to tungsten. Because it has low electrical resistivity, it finds few applications in electron tubes.

Corning 7760 has moderately high electrical resistivity, similar to that of 7720 and 7750. It contains a substantial amount of fluoride. This glass is used for the pressed funnel and neck tubing on tube types made for theater projection.

Corning 7070 has exceptionally high electrical resistivity. It is used for the faceplates of theater-projection tubes because it is the best glass to withstand electron bombardment at very high voltage. In this application it shows some X-ray browning.

Corning 3320 is used for sealing 7720 glass to 7740 glass and also for sealing directly to tungsten. This glass contains uranium oxide and has a pronounced yellow color.

Corning 7761 is a high resistivity, fluoride-free glass used for making the multiform glass beads for electron gun mounts.

MECHANICAL STRENGTH OF GLASS

A basic knowledge of the mechanical characteristics of glass is an essential requirement for those concerned with the design and processing of electron tubes.

Glass is an elastic material having a Young's Modulus (about 10^7 psi for most glasses) about one-third that of steels or most metals to which it is sealed. It is a brittle material showing practically no creep or viscous flow over extended periods of time at ordinary temperatures. It behaves as a brittle material in fracture and normally breaks in tension rather than compression. Although it has a high intrinsic strength, it must be treated as a weak material in all practical applications; for example, glass parts under continuous external load are designed for maximum tension stresses of 1000 psi, or less.

Mechanical Strength and Fatigue of Glass

When the strength of glass is measured, the values obtained are found to depend on the rate of application of stress. When the load is applied slowly, glass breaks at much lower stress values than when the load is applied rapidly. If the load is applied step-wise, with constant load applied for an interval of time, then another higher load for the same interval, and so on until fracture occurs, the glass tested with longer intervals between loadings breaks at lower stress values than the glass tested with shorter intervals between loadings. Moreover, a specimen under constant load may break several hours, or days, after the load is first applied. This weakening of glass with increased time of loading is known as fatigue. The fatigue of glass is an important phenomenon for electron tubes, as some degree of permanent strain lasting throughout the life of the tube is present in all glass members. If these strains are excessive, tube failure due to glass cracking may result in the field. Fatigue effects also characterize the failure of glass parts in manufacturing operations.

The fatigue of glass is illustrated in Fig. 1, which shows the average strength determined by cross-bending 7/32-inch-diameter glass rods, and using step-wise loading with the load intervals varied from 1/100-second to 1 day.[2,3] The glass rods, having a normal amount of light surface damage (present in all commercial glass), were tested wet. The average strengths so determined for a soda-lime-silica glass exhibit a marked fatigue effect, the average strength for a one-day load interval being only one-third of that for a 1/100-second load interval. Other glass compositions show the same behavior, with the higher silica glasses (including fused silica) showing less fatigue and slightly higher strength.

For loads sustained more than one day, a further decrease in strength could be extrapolated from the data of Fig. 1, which would indicate an average strength, for loads of one year's duration, of about 5000 pounds per square inch. As strength measurements typically have a standard deviation of more than 10 per cent, the lower 3-sigma limit corresponding to 0.3 per cent failures in one year (if a normal distribution is assumed) is less than 3500 pounds per square inch. A design strength for glass under continuous load of 1000 pounds per square inch is thus seen not to be excessively conservative, particularly when the sensitivity of the strength of glass surfaces towards damage incurred in handling and use is considered.

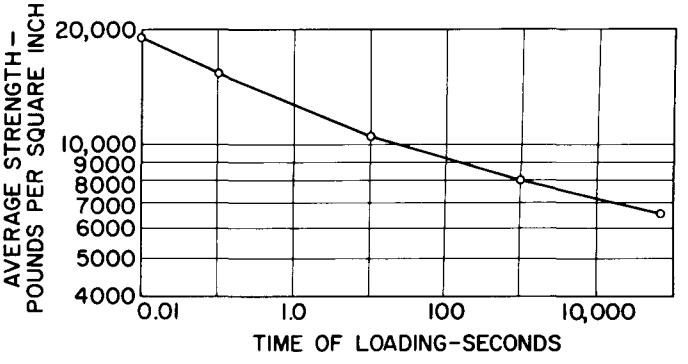


Figure 1. Breaking Strength of Glass vs. Loading Time Interval

The strength of glass is also influenced by the atmosphere in which it is tested and in particular by the humidity of the atmosphere. Glass rods tested (as described) with wet surfaces show a maximum fatigue effect which is not found in similar rods tested in a vacuum. The fatigue effect is thus thought to be a manifestation of the hydrolysis of the glass surface. (See references 1 and 4.)

Glass Fracture Analysis

When a glass article fails under stress, the crack system almost always starts at some point of imperfection on the surface of the article. Once started, the crack is propagated through the body of the article from tension stress concentrations formed at the point of the crack. If the glass is not highly stressed at the time of failure, the crack may propagate as a single crack, varying its direction, but stopping only when the article falls apart or when the applied stress is sufficiently relieved by the extension of the crack. If the glass is highly stressed at the time of failure, the crack may propagate in a complicated way, developing a forked structure and ending up with the article split into small pieces of glass. The complexity of the crack pattern and the average size of fragments left after failure indicate, to some degree, the energy of the stress system which caused failure. The point at which failure first occurred can be found by visual examination of these fragments. By observation of the ripple pattern of the fracture surfaces (see Fig. 2), the direction of motion of the crack can be determined.[5] The glass piece containing the origin can be located by assemblying the glass pieces and tracing out the development of the crack system. The origin itself is frequently (but not always) well-defined by the circular ripple marks which surround it.

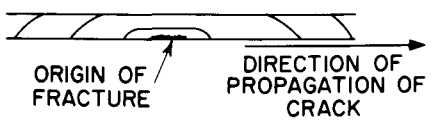


Figure 2. Ripple Pattern on Glass Fracture Surface

In the analysis of a failure, an obvious surface flaw or defect is usually found to be the origin of the fracture. This defect is often damage to the surface, such as a scratch or bruise mark. A scratch may have a small invisible crack at its base, and under sufficient tension stress a fracture origin coincident with a small section of the scratch may cause failure. A scuff mark left by rubbing or striking the glass surface may consist of parallel scratches and, in addition, be bruised and exhibit small cracks or checks at right angles to the scratches. The checks are crescent-shaped and are due to the glass surface mechanically seizing the object in contact with it.

Checks, unlike scratches, are frequently invisible but, when present, may readily be found by the use of a light etch of the surface with hydrofluoric acid and examination of the etched surface with the aid of a low-power microscope. Checks are among the most serious weaknesses in a glass surface. Under sufficient tension stress, a check readily develops into a fracture origin, the initial direction of the resulting crack being at right angles to the direction of scratches in the bruise mark. Checks and scratches provide characteristic fracture origins which are readily diagnosed by visual examination.

Clean glass surfaces are most subject to damage by transient contacts with other materials. The problem is greatest during a cooling operation, before the glass surface has collected a lubricating film of water and oil from the air. Sometimes silicone compounds are applied after bake or annealing to provide lubrication and minimize this kind of surface damage. Although effective, this treatment is not desirable for glassware used in electron tubes because the compounds used form contaminants which are harmful and difficult to remove.

Other surface defects or imperfections are associated with the original fabrication of the glass article. These include laps, tear marks, and shear marks which are found in pressed or molded glass articles. These defects may form fracture origins and lead to failure of glass pieces in some operations or applications.

The interface of glass-to-metal seals is another common origin of fractures in electron-tube applications. These fractures may occur when the glass-to-metal interface adherence or the galss geometry is poor—conditions which are worsened by strains at the interface due to the mismatch in thermal expansion characteristics of glass and metal. These origins are among the most difficult to classify and treat in production.

In a practical shrinkage or breakage problem, the first step is to locate typical origins and determine the type of surface flaw at which the origins occur. If the flaw is a manufacturing defect in the glass article, it may be possible for the glass manufacturer to improve the defect, or relocate it (e.g., if the flaw is a shear mark) to an area of the article which is not so highly stressed. If the flaw is a scratch or bruise mark, it may be possible to modify the processing procedures which are responsible for this surface damage. If the flaw is one representative of commercial glass quality which cannot reasonably be improved, then resort must be made to other possibilities for decreasing breakage such as reducing tension stresses imposed on the defect in the operation or application, or inducing controlled compressive stresses in the critical surface areas. The methods described in the next section for estimating tension stress may be applied to determine whether excessively high stresses are present.

Measurement of Stresses in Glass

Stresses in the glass parts of a tube envelope may be classified into four groups:

1. Stresses near glass-to-metal or glass-to-glass seals due to differences in thermal expansion characteristics of the two components of the seal assembly
2. Stresses associated with incomplete annealing of the seal assembly

3. Stresses due to nonhomogeneities in the glass (i.e., cords, streaks, layers of glass of varying composition, stones, or foreign inclusions)
4. Stresses due to external conditions (such as external atmospheric pressure, impact, mechanical loading, or nonuniform heating)

Stresses may be measured or estimated by the use of a polariscope, strain gages, or similar devices. Such measurements, together with strength measurements carried out on the glass article, are the basis for the evaluation of design and processing parameters.

Reference 6 gives a general treatment of stresses in thin-walled symmetrical bodies, such as are usual to tube envelopes, and will supply useful background information on these stress systems.

Polariscopic Estimation of Stress. The polariscope* is the most commonly used means for observing stresses in glass pieces. It may be used quantitatively if the stresses have a simple and uniform distribution but the complicated stresses usually encountered can only be qualitatively evaluated. The production practice is usually to establish a desirable stress pattern empirically using the polariscope, and then to control the annealing of the glass part to maintain this pattern.

The polariscope uses polarized light to measure the retardation or phase shift of light of one plane of polarization relative to that polarized in the perpendicular direction, as it passes through the glass under test.[7] Retardation or phase shift may be given in degrees of rotation or of fractions of a wave length and it is usually measured with green light of wave length of 540 millimicrons and expressed in millimicrons. For additional information on the optical theory of the polariscope and its applications to stress measurements see reference.[8] Retardation observed in stressed glass is usually a small fraction of a wave length; the polariscope is commonly used with either a full-wave or a quarter-wave retardation plate. These are elements which produce the specified retardation at a wave length of 540 millimicrons. Three difference arrangements for the use of a polariscope with glass samples are shown in Fig. 3.

The full-wave plate (or tint plate) is placed between crossed polarizing elements in the orientation to produce its maximum retardation (Fig. 3A). With a white-light source, the dark field due to the crossed polarizing elements thus becomes a first-order magenta color (the 540 millimicron green is absent, but some red and blue get through). A glass test sample placed in this instrument shifts the extinction color from green towards either red or blue; the appearance of the opposite color (blue or red) is a sensitive indication of retardation in the sample. The intensity and amount of color shift can be used, by comparison with standard strain samples, to estimate retardations up to about 100 millimicrons; beyond this value the color changes are less pronounced. The sign of retardation indicated by the direction of the color shift (red vs. blue) may readily be related to the sign of stress (tension vs. compression) by the use of a strip or rod of glass, which is stressed by bending with the fingers. The full-wave polariscope gives a convenient pattern of color, corresponding to retardation, which can be seen over the entire glass piece.

A more quantitative measurement of retardation is possible with the quarter-wave plate. This element is placed between crossed polarizing elements in the orientation to give zero retardation (Fig. 3B); it then corrects the elliptical polarization of light which has passed through the glass sample, restoring it to plane-polarized light, but with the plane of polarization rotated. This angle is determined by rotating the polarizing element to obtain extinction. Retardation is twice the angular rotation of the polarizing element. Thus, an angle rotation of 1 degree represents a retardation of 2 degrees or of (2/360) x 540 = 3 millimicrons. This polariscope may be used with a white-light source for retardations of less than one wave length; beyond this value a monochromatic green light source may be preferred. The quarter-wave polariscope indicates retardation at a locus or fringe, the fringe being moved by rotation of

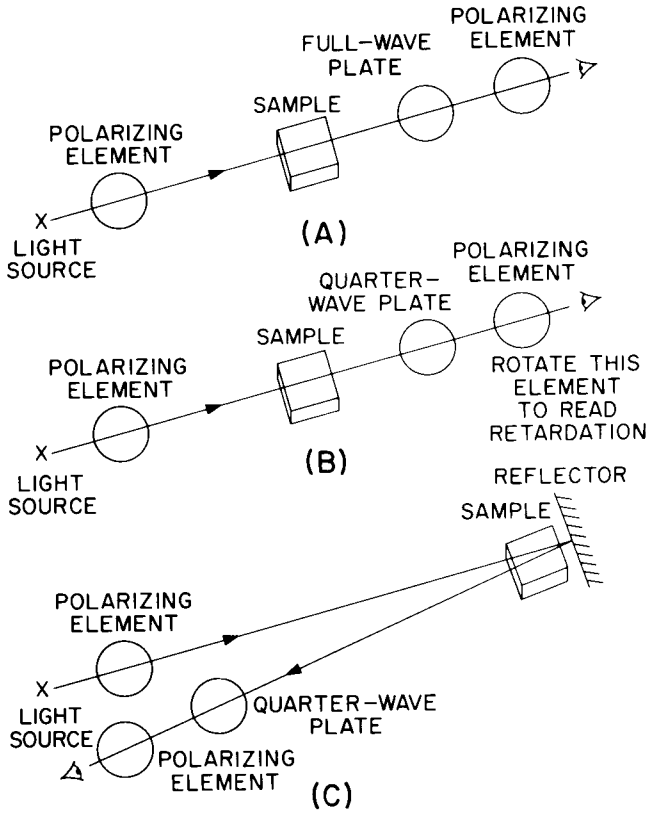


Figure 3. Three Arrangements for Polariscope Use with Glass Samples: (A) full-wave retardation (tint) plate, (B) quarter-wave retardation plate (Friedel polarimeter), (C) reflection arrangement

* The instrument used for quantitative measurement of retardation should be called a polarimeter; however, according to common usage, it is referred to here as a polariscope.

the polarizing element to coincide with a point of interest on the glass sample, and thus gives a point-by-point quantitative measurement of retardation rather than an over-all pattern such as the full-wave polariscope displays.

The polariscope is usually used with light transmitted only once through the glass object. For thin-walled articles, however, the transmitted light may be reflected back through the article (Fig. 3C). With this arrangement the light path and the observed retardation are doubled. This kind of device has been used for observing retardation patterns in large kinescope funnels after exhaust processing of the tube.

Several types of one-dimensional stress relations are susceptible to investigation with the polariscope. Retardation observed in a glass piece which is uniformly stressed in one direction is proportional to both the thickness of glass and the stress (Fig. 4A). Retardation expressed in millimicrons per centimeter of thickness (or optical path length) is, in this case, a measure of stress. Remember that a stress of 1 kilogram per square millimeter (1400 psi) corresponds roughly to a retardation of 250 millimicrons per centimeter in the high-expansion glasses, or 350 millimicrons per centimeter in the low-expansion glasses. This is a rough limiting value for tensile stress which should exist as a permanent stress at the surface of small glass pieces. (For stress optical coefficients, see Table III).

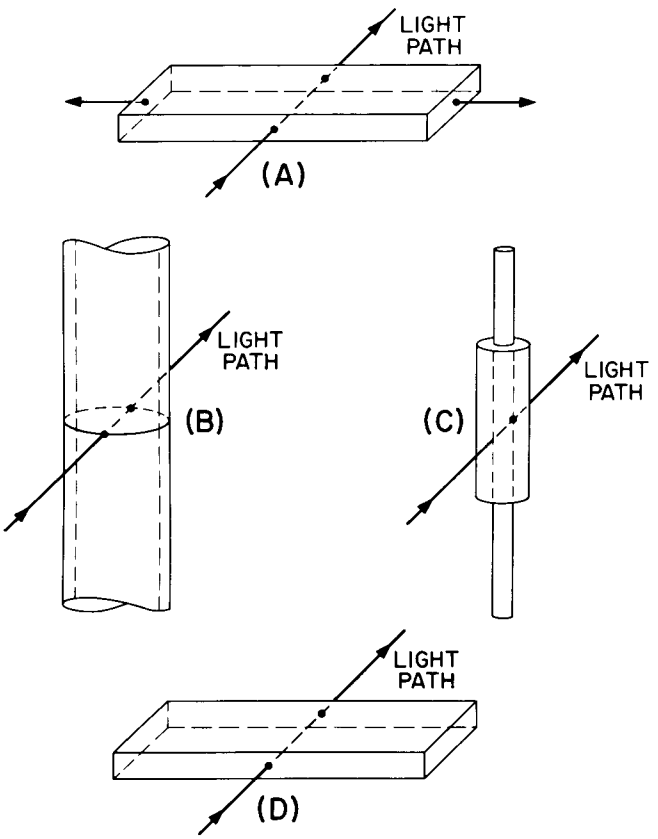


Figure 4. Examples of One-Dimensional Stress-Retardation Relation Which Can Be Determined in the Polariscope: (A) glass piece with a uniform one-dimensional stress, (B) butt seal of glass tubing, (C) bead sealed to wire, (D) tempered flat glass piece

Table III

Stress-Optical Coefficients of Glasses

Type of Glass	Stress-Optical Coefficient (μ/cm/kg/cm^2)*
Silica	3.4
Borosilicate, low expansion	3.8
Aluminosilicate	2.6
Soda-lime-silica	2.4-2.6
Lead-alkali-silica	2.5-2.7

*The coefficient is given in millimicrons retardation per centimeter of optical path for an applied stress of one kilogram per square centimeter.

In a few practical cases, retardation can be interpreted in terms of a one-dimensional stress. The stresses in a butt splice of tubing of two different glasses, in a butt seal of glass tubing to metal, or in the "ring stain" left in glass tubing after sealing—all have a similar distribution consisting of longitudinal bending stresses and tangential or circular hoop stresses. If the glass tubing is viewed centrally perpendicular to its axis (Fig. 4B), retardation of the bending stresses integrates to zero and only retardation of the hoop stresses is seen. The hoop stress can thus be calculated as shown earlier, using twice the thickness of the tubing as the optical path length.

The stresses in a long bead of glass sealed to a metal wire can be similarly interpreted. If the bead is viewed perpendicularly to its axis (Fig. 4C), the retardation of the radial and tangential stresses integrates to zero, and only retardation produced by the axial stress is seen.[9] Axial stress is determined from retardation measured next to the wire edge, with the length of light path given by $2\,(b^2-a^2)^{1/2}$, i.e., that of the ray which is tangential to the wire edge.

A one-dimensional stress pattern is seen when a strip of tempered flat glass is viewed edgewise (Fig. 4D). In this case, stresses are parallel to the flat surface, with tension at the center and compression at the edges of the strip; their values can be obtained by calculation from the retardation pattern. The maximum central tension stress is used as a measure of the degree of tempering of the glass. The same kind of examination can be made on a flat piece of glass cut or sawed from a thin-walled glass piece. In this case, the center tension is sometimes used as a convenient measure of the effects of rapid annealing schedules occasionally used in production.

Retardation can be seen in the walls of tubing or bulbs by immersing the part in a cell containing an index-

matching liquid* and viewing the part tangentially. The cross section of the wall seen in this way displays a complex retardation pattern not readily interpreted in terms of stress. In some cases, however, a useful indication of bending stresses due to metal seals or of longitudinal stress due to rapid annealing is possible and has been used for production control purposes. (For a mathematical treatment of this problem, see reference 10.)

Button stems are frequently examined by viewing them axially in the polariscope with or without immersion in an index liquid. The retardation pattern is complex and, therefore, stress analysis is not usually attempted; however, for production control, requirements may be established for maximum or minimum "tension" or "compression" readings at different points of the stem. Glass stems with metal tubulation can be examined for retardation at the metal seal, and, based on the readings, control of the sealing and annealing schedule is possible. A similar type of control is used for top seals on small bulbs, where readings are obtained on the part viewed along the bulb axis. A layer strain can be seen in button stems by viewing them from the side in an immersion liquid; however, this method is seldom used for production control.

Nonhomogeneities (cords) in glass articles are caused by poor mixing of the glass during melting and are usually present in layers drawn out parallel to the surface of the article during the manufacturing process. Retardation of the layers tends to integrate to zero when the piece is viewed perpendicularly to the wall. The layers are best seen by cutting the article to obtain a cross section which can be viewed in the direction of the glass flow. Evaluation of their intensity is difficult. If large areas or layers approaching the limiting value of 1400 psi are present, there may be an adverse effect of the nonhomogeneity on strength of the glass article. This effect on strength is better detected by a thermal shock test than by other types of strength test.

Measurement of Strains in Glass Surfaces. The polariscope measures an integrated effect of the volume or body stresses and thus only indirectly indicates surface stresses, which are frequently of greater interest than the body stresses. For a quantitative measurement of strains at the surfaces of glass articles, the bonded wire-strain gage can be used.

The SR-4 strain gage is a fine wire grid, usually wound on and cemented to paper, which is in turn firmly cemented to the glass surface. When the glass article is externally loaded, the surface strain is transmitted to the wire gage, whose electrical resistance varies rapidly with elongation or strain. The direction and magnitude of the principal stresses at the surface are readily obtained by means of these gages.

The type A-7 gage ** is paper mounted; its size (1/8 x 3/16 inch) makes it convenient to use; smaller gages are also available. An epoxy adhesive such as Armstrong's A-1 cement is used for mounting the gage. Changes in gage resistance can be measured with commercial instruments, or on a simple Wheatstone bridge circuit; with ordinary techniques, an accuracy of measurements of ± 10 microinches per inch can be expected.

The use of strain gages normally is limited to determining the strains associated with an external load at room temperature. The method is used in checking strains set up in large kinescope bulbs during exhaust. The change in gage resistance is determined when air is pumped out of the bulb. By this method, the areas of maximum tension stress may be found so that design parameters of the bulb can be established to insure adequate strength. Large glass kinescope bulbs are designed for a maximum tensile strain of about 100 microinches per inch (a value which corresponds to a tensile stress of 1000 psi).

In some cases, strains imposed on the glass surface by the metal part of a glass-to-metal seal may be determined with a strain gage when the glass member is broken away from the metal member. Care must be taken that the glass under the gage does not crack. An alternate method would be to dissolve the metal member away with an acid. A similar technique has been used to determine "built-in" strains due to incomplete annealing of large glass pieces; in these cases, after the gage is cemented on the article, the article is broken up or sawed apart so that the gage measures the strain in a relatively small piece of glass.

Accurate methods for using bonded strain gages at elevated temperatures would be of considerable interest for solving problems encountered in the processing of glass articles. However, because the strains of interest are relatively small compared with the interfering effects of temperature change, useful measurements of strains during heat treatment are difficult.

Other methods are available for determining strains in surfaces. A brittle lacquer† may be used to locate highly strained areas and to characterize the principal stresses in these areas. In this method, the surface is strained after application of the lacquer and the crack system which the lacquer develops is observed. However, the strains in glasses are normally at about the lower limit of sensitivity of this material.

A polariscopic method‡ of determining surface strains uses a birefringent coating applied over a metallic reflecting coating on the surface; the two-dimensional retardation pattern of the coating is then observed when the surface is strained. Either of these methods may be preferred to wire strain gages in cases where complex glass shapes and strain patterns are involved.

Scratch Testing of Stresses in Glass. The scratch test is a useful (but destructive) practical test of glass

*Suitable liquids are trichloroethylene (index 1.48) for borosilicate glasses, and o-dichlorobenzene (index 1.55) for lead glasses.

** Baldwin-Lima-Hamilton Corp.

† Stresscoat, Mangaflux Corp., Chicago, Ill.

‡ Photostress, The Budd Company, Philadelphia, Pa.

stresses of any origin; in this test, the glass surface is purposely scratched or abraded to weaken it. It is useful where glass breakage is being experienced at some production operation. If the abraded glass article consistently breaks in such an operation, unduly high stresses are being imposed on the glass at some stage of processing. The origins so obtained indicate the part of the surface which is receiving the most stress. If the abraded part fails to show increased breakage during the operation, it may be assumed that there is some other cause for the glass failures than high stress; probably the cause is a very weak defect somewhere in the glass surface.

This test may also be used to determine whether manufactured glass pieces have stresses sufficiently high to cause fatigue failure in the field. If the article fails (even slowly) after being abraded, it is usually safe to assume that some field trouble will be experienced.

Estimation of Stresses by Strength Testing of Glass Articles. Glass pieces may be strength-tested to evaluate their probable performance through processing or in final use in electron tubes. Among the methods used are thermal shock testing, mechanical testing, and pressure testing. In these tests, an attempt is made to duplicate and to exceed the stresses normally encountered so that breakage of weak ware will be increased.

For example, bulbs for cathode-ray tubes must withstand atmospheric pressure (15 psi) throughout their life without collapsing. Glass bulbs are commonly tested at 45 to 60 psi in incoming lots. However, this is not really a severe test for new bulbs which have not yet received any surface damage. Consequently, an abraded test is preferred in which the critical areas are abraded with 80- to 100-grid abrasive paper before the pressure test, which is made with the glass surfaces wet. Experience has shown that large glass bulbs should pass such a test at an external pressure of about 45 psi.

Thermal shock tests are used particularly for testing seals and sometimes in conjunction with a mechanical test. Mechanical testing of seals is a good method for determining the seal geometry or processing that will give the highest strength.

ELECTRICAL PROPERTIES OF GLASS

The performance of glass as an electrical insulator is characterized by physical properties of the glass, such as volume resistivity, power factor, and dielectric constant. Tabulated values of volume resistivity were given earlier; however, power factor is a complicated function of frequency and temperature and no simple tabulation of this property is possible.

The electrical resistivity of glass increases exponentially with temperature. If log resistivity is plotted against the reciprocal of absolute temperature, a slightly curved line is obtained. Because the loss factor of glass depends strongly both on temperature and on frequency, the effects of these parameters should be considered in estimating the values of electrical properties for a particular glass application.

Electrical properties also depend on the previous history of a glass. Current conduction is accompanied by strong polarization; as a result, the resistivity values generally given are obtained at low electric fields and low currents to minimize the polarization effect. Previous heat treatment has a strong effect; glass has higher losses and lower resistivity if it has been rapidly cooled through the annealing range than if it has been slowly cooled. Therefore, the values of electrical properties of commercial glasses are usually given for well-annealed samples.

Electrical Failure of Glass

In electron-tube applications, glasses frequently operate at elevated temperatures. Although conditions causing true dielectric failure are probably never encountered, there is a possibility of electrical failure in service due to thermal breakdown under high electric fields and high temperature. This condition may limit the use of glass compositions for some applications to those of higher resistivity or lower loss. In some high-frequency power tubes, because no commercial glasses meet all the requirements of tube operation, it is necessary to use other ceramic materials whose electrical properties are less sensitive to temperature.

Electrolysis Effects. The conduction of current in glass is ionic, that is, the charge carriers for the conduction current are ions rather than electrons. Glass, therefore, shows electrolysis effects which are due to the transportation of matter associated with the conduction current. Thus, if dc voltage is applied across two leads sealed into glass, the mobile ion concentration will decrease in the glass surrounding one lead, and will increase around the other lead. If the conduction current is large, electrolysis will produce substantial volume and stress changes with eventual failure of the glass by cracking. An inverse measure of the effect is provided by the volume resistivity of the glass.

Electrolysis effects are most likely to occur in the stems of small tubes, where temperatures are moderately high, leads are close together, and appreciable voltages are applied between the leads. For this reason, stems are made with glasses of high electrical resistivity. With 0120 glass, electrolysis effects are usually negligible at stem temperatures below about 150 C. For stems operating above this temperature, it is advisable to suspect electrolysis; the best indications of it are changes which occur in the lead color during normal tube operation. Another method is to observe changes in the polariscopic strain pattern in the glass around leads. If these effects are seen during early operation of a tube, there is a possibility that the tube will fail from glass cracks due to electrolysis during extended life operation.

The sodium ion is the charge-carrier for conduction of current in soda-lime-silica glasses, and presumably also in most other commercial glasses.* During the

*There is little literature on the identity and behavior of the charge-carrying ions in commercial glasses.

conduction of current in glass, sodium metal is formed at the negative lead or electrode, and oxygen gas at the positive electrode; more generally, there is chemical reduction at the cathode and oxidation at the anode leads. Thus, in a lead glass, although Na^+ is the mobile ion, metallic lead is produced at the negative lead wire. The lead deposit causes a prominent darkening of the lead wire which appears first as dark spots and then grows to a dendritic or tree-like form. Darkening of the anode lead wire due to oxidation can be observed as the first effect of electrolysis; subsequently, the effect may occur in the form of spots which grow larger and darker. It seems doubtful that electrolysis produces any release of oxygen gas into the tube at the anode lead-- at least, no quantitative data on this phenomenon are known.

Electrolysis effects in some tube types are complicated by the bombardment of glass surfaces by electrons escaping from the mount, in which case the current path may be between the bombarded area and one or more positive leads. In high-voltage tubes, these effects can be erratic because the bombarded area may either charge up nearly to the anode potential, or it may have its potential suppressed to a value close to that of the emitter.

Puncture. Puncture of the glass portion of a tube envelope occurs when a small area has been heated by electrical losses until the glass melts and sucks in or cracks. It results from a runaway condition in which electrical losses, in glass heated initially by the conduction current or by dielectric loss, increase with increasing temperature until a catastrophic hot spot develops. A characteristic of puncture, or cracking by incipient puncture, is that localized permanent strains (caused by excessive heating and visible in a polariscope) are introduced in the glass close to the point of failure. This effect may be found typically in the neck and funnel portions of high-voltage projection tubes, where electrical fields of the order of 200 to 1000 volts per mil may be applied across the wall section, and temperatures of 70 C, or more, may be produced by heating from the yoke which surrounds the neck and lower funnel of the tube. Under these conditions, random failures during the first few hundred hours of tube operation may be observed in glasses of medium electrical resistivity.

Bulb punctures may similarly be observed in high-frequency power tubes, where the glass is heated by a combination of factors including dissipation losses in elements, resistance losses in the metal surfaces of the glass-to-metal seals, and dielectric heating of the glass itself. The possibility of bulb puncture is a real limitation on the use of glass for the envelopes of high-frequency, high-power tubes.

Puncture or cracking may be due to electron bombardment of the bulb in some high-voltage tubes. Failure usually results from intense bombardment of a localized area and may result in a star-shaped crack system or a suck-in. This condition is observed in regions of high electric fields, i.e., on the surface of the bulb closest to the anode of a high-voltage rectifier tube, or adjacent to the high-voltage elements of a gun. With such tube types, it may be helpful to cover affected areas with an aluminum-oxide coating made with a sodium silicate binder. This coating provides some slight conductivity and has higher secondary emission than glass. Early evidence of bulb bombardment is provided by fluorescence of the bombarded areas and in some cases by browning of the glass in these areas.

Ionization in Air Lines and Bubbles. Bubbles in glass* contain gases at less than atmospheric pressure. When subjected to high electric fields, these defects sometimes cause glass failure due to heating by ionization of the gases. Early indications of ionization are the lighting-up of air lines or bubbles during high-voltage operation.

In some applications, it is necessary to specify an upper limit to the size of bubbles or the diameter or length of air lines. As an example, the neck tubing of some high-voltage cathode-ray types is required to have no air lines of diameter greater than 0.001 inch near the inside surface in the area close to the high-voltage part of the gun. Air lines in the cane glass used for gun mounts may have to be limited similarly because of failures during high-voltage spot knocking. For gun mounts, multiform glass (made by sintering glass powder) can be used because the bubbles in it are very small.

Conductive Films on Glass Surfaces

In many applications of glass to electron tubes, the use of glass with normally high surface-resistivity is necessary for proper functioning of the tube. Electrical leakage on external surfaces, from weathering or absorption of water from humid air (relative humidity of over 50 per cent), may be overcome by application of coatings such as silicone films or water-resistant paints. Electrical leakage on internal surfaces may be caused by contamination before or during exhaust processing, e.g., by deposits of carbon compounds or volatile metals such as silver or copper. Low grid resistance may sometimes be found due to internal leakage in tubes where the glass surfaces show no visible contaminating films.

In other applications where some electrical conductivity of glass surfaces may be required, opaque conductive coatings such as an evaporated aluminum film, platinum or silver coating, or silicate-bonded graphite coating are usually used.

Transparent, iridescent, electrically conducting films can be formed on glass, mica, or porcelain surfaces by exposing them while hot to fumes or sprayed solutions of tin compounds. The process has been called "iridizing;" the films are referred to as "EC" coating by Corning Glass Works, as "TIC" coating by RCA, and "NESA" coating by Pittsburgh Plate Glass Company. The object after coating has a uniformly thin

* Bubbles in drawn tubing (frequently called air lines) are elongated and thus are particularly susceptible to ionization effects.

film of tin oxide closely adherent to the glass surface. Thickness of the film may vary from less than 100 millimicrons to more than 1000 millimicrons, and surface-coating resistance may vary from less than 1000 ohms per unit square to more than 1 megohm per unit square. A surface coating resistance of 5000 to 10,000 ohms per unit square is practical in films which reduce light transmittance by no more than 10 to 15 per cent.

A typical application involves spraying a heated glass surface with a solution containing stannic chloride which might contain, for example, 5 parts of $SnC1_4 \cdot 5H_2O$ and 1 part of anhydrous methanol. The glass temperature should be close to the softening point at the time of spraying. The use of conducting films based on other metals has been reported; these make use of compounds of cadmium, zinc, or indium.

OPTICAL PROPERTIES OF GLASS

In electron-tube applications, glass has few critical requirements such as are met in the field of glass for optical instruments, which generally require pieces of highly homogeneous glasses having accurately finished surfaces. Special requirements for tubes usually involve only the controlled transmittance or absorption of light or radiation. Faceplates of optical surface quality are used in some pickup and phototubes, however, and melts of optical quality glass are occasionally used to obtain glass pieces of exceptional homogeneity and freedom from melting defects.

Transmission of Light and Radiation

Commercial glasses are transparent to light and to radiant energy in the adjacent parts of the infrared and ultraviolet spectrums. The range of transmission for the best silicate glasses (excluding fused silica) is from a wave length of about 200 millimicrons to 2500 millimicrons with reduced transmission to almost 5000 millimicrons.

Phototubes, image tubes, and similar tube types using photocathodes require a bulb or window glass having good transmission for visible radiation and, for some applications, either good ultraviolet or infrared transmission is also required. Glasses 0080, 7052, and 9741 are generally used; transmittance for these glasses is shown in Fig. 5. Glasses 0080, 7052, and 7056 have ultraviolet transmittance down to a wave length of about 300 millimicrons; glass 9741 provides useful transmittance down to a wave length of about 200 millimicrons; a fused silica window is transparent to radiation at a wave length below 180 millimicrons. The silica window bulbs are made with a graded seal so that a 7052 glass stem with kovar leads can be used for the mount.

Kinescopes are commonly made with a gray-glass faceplate which absorbs 40 to 70 per cent of the reflected ambient light to improve the large-area contrast of the picture.[11] Transmittance of a typical kinescope glass (Corning 9010) is shown in Fig. 6. Absorption is due to the presence of small quantities of cobalt and nickel oxides in the glass; the amounts are adjusted to give the glass a nearly neutral tint. The color specification of these glasses requires that the ICI coordinates of a standard light source employing a P4 phosphor, measured after transmission of the light through the faceplate, should not shift outside the boundaries of a small polygon of specific size.

All commercial glasses show infrared absorption above the water absorption band at a wave length of 2.8 microns, and absorption is complete for wave lengths above 4 to 5 microns (see Fig. 7). Although no electron tubes require glasses of exceptional infrared transmittance, some devices (such as photocells) manufactured at Lancaster do. For these applications, materials other than silicate glasses are required for windows.

As an envelope material, glass assists in the cooling

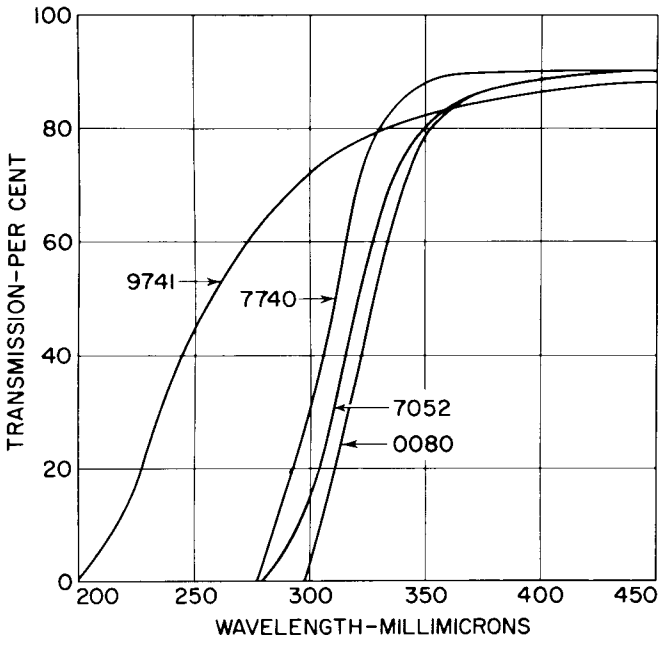


Figure 5. Ultraviolet Transmission of Some Common Glasses (2-mm sample-thickness)

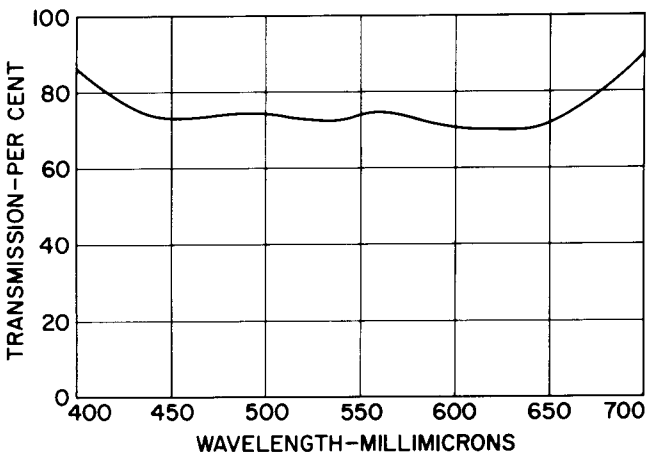


Figure 6. Transmission of Corning 9019 Gray Glass

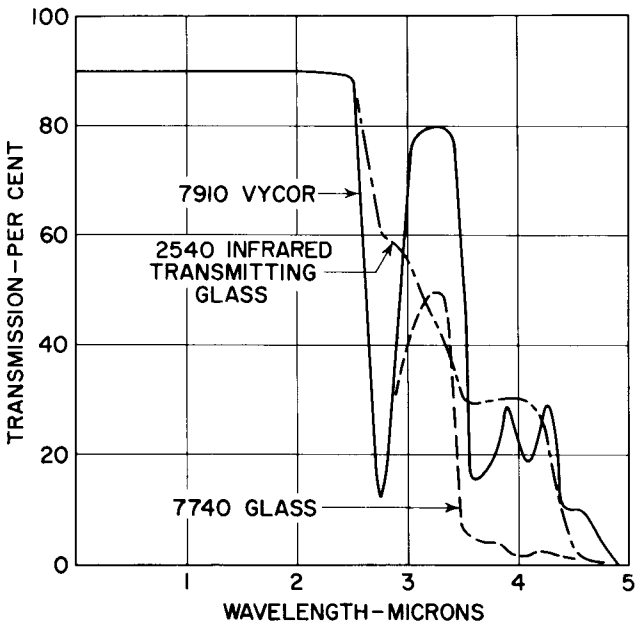


Figure 7. Infrared Transmission of Some Glasses

of internal tube parts because it does not reflect infrared radiation of any wave length. Most commercial glasses (see Fig. 7) transmit a high percentage of the heat radiation from bodies at temperatures above 1000 C (reference 1, page 66) while the heat radiation of longer wave length from bodies below this temperature is efficiently absorbed by glass. However, hot glass heated by energy absorption does radiate heat internally as well as externally. The result is an increase in ambient temperature within the tube and some increase in element temperature.

X-Ray Absorption

Efficient absorption of X-rays by the envelope of some high-voltage electron tubes (excepting X-ray tubes) is desirable to control radiation levels near operating equipment. Radiation is no problem for tubes operating under 15 to 20 kilovolts because the soft X-rays produced at these voltages are readily absorbed by the usual glass envelopes. Above 20 to 25 kilovolts, however, glass envelopes of suitable thickness and absorption characteristics may be required; for such envelopes, lead or barium glasses may be used to provide high absorption.

The absorption of X-rays by glasses is calculated by using a linear equation involving the mass absorption coefficients of the oxide components of the glass and their concentrations (see ref. 1, p. 82).

Glass Browning

An interesting property of glasses is the development of absorption bands on exposure to radiation or electron bombardment. These bands are caused by displacement of electrons from their normal positions, to form color centers which absorb light in a characteristic range of wave lengths, mainly those in the ultraviolet region. For many glasses, one of the bands overlaps the visible spectrum so that the glass colors and darkens on exposure; the color usually is a yellow or brown, although some glasses develop a blue color; this phenomenon is called browning.

Two browning effects occur when glass is bombarded by high-voltage electrons. One kind is browning caused by electron penetration of the glass, an effect necessarily confined to a thin surface layer. The other kind of browning is caused by X-rays generated by electron bombardment and is distributed through a much larger section of the glass wall. The form in which browning takes place may be used to tell whether it is due to electron bombardment or X-ray penetration.

These effects may be objectionable in some cathode-ray types where the glass picture area is subject to bombardment by high-voltage electrons. During operation at screen voltages below 15 kilovolts, browning is not found with commercial glasses; above 20 to 25 kilovolts, however, glasses of special compositions are used to reduce browning effects that develop during extended tube operation.

Most glasses show some X-ray browning, but the amount can be reduced by small additions to the glass composition; visible effects can be greatly reduced by additions of oxides of transition elements such as manganese, iron, cobalt, nickel, and vanadium.[12, 13] Cerium oxide has a similar effect and is desirable because it does not color the glass as do the foregoing oxides; additions of 0.3 to 1 per cent of cerium oxide are present in the panel glasses used for color kinescopes (Pittsburgh Plate No. 5533 and Corning 9019 glasses), as well as in an earlier glass (Pittsburgh Plate No. 3459 glass) used for a 5-inch projection tube.

X-ray browning fades slowly with time at room temperature and can be accelerated by increasing the temperature. Heating glasses to a temperature between 200 and 300 C removes X-ray browning.

Most glasses also show electron browning, which is quite permanent and cannot be removed by heat treatment. Electron browning is not appreciably reduced by the addition to the glass compositions of the transition elements which reduce X-ray browning. However, some compositions such as Pittsburgh Plate 3459, Corning 9010, and Corning 7070 borosilicate glasses show insignificant amounts of electron browning in cathode-ray-tube applications.

Sometimes a severe browning effect develops on the stem or bulb during the life of some tubes operating at relatively low voltages, particularly with some of the older designs of small power-tube types. The cause is believed by some to be electron bombardment of the affected areas, but a more likely explanation is that the browning is due to electrolytic reduction of lead rather than to electron penetration, because the same types also show the typical darkening around stem leads which is associated with electrolysis.

Fluorescence of Glass

Glasses generally fluoresce when exposed to ultra-

violet radiation or to high-voltage electron bombardment. Neither phenomenon has any practical application because the effects are so weak, but may be useful in identifying glasses of unknown composition. After being irradiated, and when it is heated to remove the browning effect, glass also shows thermoluminescence.[12,13] The light emitted during heating to remove browning is a more sensitive indication of radiation effects than is browning itself; thermoluminescence, in fact, is being applied to the dating of very old (100, 000 years) ceramic pieces.

Homogeneity — Glasses Melted for Optical Use

Because of their relatively poorer homogeneity, most commercial glasses are usually not suitable for use in lenses, prisms, or other refractive elements of optical systems. Optical glasses are specially melted to attain homogeneity and, since 1945, continuous melting methods have been developed for these glasses. Many critical glass pieces are made in this manner for electron-tube applications. These pieces include small faceplates for camera tubes, phototubes, image tubes, and similar types which use glass in the form of round flat sections or lenses with ground and polished faces. These needs are supplied in Corning 7056 kovar-sealing glass. Another application for such glass is the free-blown bulb blank used for the thin glass target of the image orthicon. Bausch and Lomb Company supplied the first optical glass used at Lancaster — a BSC-51 glass faceplate used on the image orthicon and made by spectacle-glass manufacturing techniques.

OUTGASSING, DIFFUSION, AND SUBLIMATION OF GLASS

Glass is permeable to some gases and holds these as well as some other gases in solution. Gases evolved by glass depend on the heating conditions; on heating glass, water first evolves from the hydrated surface film and, as the glass reaches somewhat higher temperature, more water diffuses out of a thin surface layer of the glass. Further heating to very high temperatures evolves carbon dioxide, sulfur dioxide, oxygen, and water. During electron bombardment, oxygen is the predominant gas evolved, with accompanying dissociation of the glass structure. Glass is permeable to helium and, at high temperatures, to hydrogen, deuterium, and neon; but these gases have a limited solubility in glass and, therefore, their evolution by glass surfaces in most practical situations is unimportant. Although little is known about the effects of ion bombardment of glass surfaces, such treatment is known to be an effective method of cleaning up glass surfaces at moderate temperatures. Copious amounts of gases such as carbon monoxide, nitrogen, nitric oxide, oxygen, carbon dioxide, and hydrocarbons can be released by sparking glass surfaces before bake-out. The liberation of gases by glass during electrolytic conduction has had relatively little attention. Finally, the sublimation of compounds from glass surfaces, even at the moderate temperatures used in electron-tube exhaust, is known to be (in some instances) an important source of oxide-cathode contamination. More information, however, on all of these phenomena would be helpful to the electron-tube industry.

Thermal Outgassing of Glass

When glass is heated, it evolves gases which are primarily water vapor.[14,15] The volume of water V evolved from a unit surface at constant temperatures above 300 C can be obtained from the equation:

$$V = m\, t^{1/2} + s$$

where t is the time of bake-out. The intercept s, which may be altered by different surface treatments, is a measure of the easily removed water residing at the surface. The slope m is a measure of the rate of evolution of water diffused to the surface from the interior. If the initial concentration of water is C_0 and the diffusion constant is D, then the slope m is given by

$$m = 1.13\, C_0 D^{1/2}$$

If m is plotted on a logarithmic scale against the reciprocal of absolute temperature, a straight line is obtained. The activation energy for the diffusion process obtained from this plot varies for usual glasses from about 35 to 50 kilocalories per mole.

The magnitude of the diffusion constant of water in glasses determined in this way shows that, during the usual exhaust bake-out conditions, only a thin layer of glass near the surface is involved in the diffusion process. Typical values of the thickness of this layer are less than about 20 microns.

If m is known as a function of temperature for a glass with initial uniform water concentration, the amount of water which will diffuse out with time at any temperature after an arbitrary bake-out can be calculated. This information can be applied to evaluate the relative effectiveness of multistep exhaust-bake processing. It can also be applied, by wide extrapolation, to evaluate the expected diffusion of water into the tube during normal tube operation. Such an extrapolation predicts that a minute and usually negligible amount of water will diffuse into tubes operating at temperatures below 150 to 200 C. At higher operating temperatures, the possibility of substantial diffusion of water into the operating tube would have to be investigated in terms of the particular gases involved and a specific exhaust process.

Corning 0080 glass was found to be in equilibrium with water at a vapor pressure of 10 to 12 millimeters of mercury at temperatures near 500 C.[14,15] Thus, at higher vapor pressures, water will diffuse into 0080 glass, while at lower vapor pressures it will diffuse out. It was evident that this glass can be outgassed by diffusion by heating it in a dry atmosphere.

The amount of surface water indicated by the value of the intercept s is rapidly and completely removed at exhaust-bake-temperatures above 300 C. The amount of water released by the glass surface increased with the age of the glass and is greatly increased if the glass is stored in a high-humidity atmosphere at 90 C. The source of this water is probably a hydrated surface layer. This layer can be removed by a mild hydrofluoric-acid wash which reduces the surface gas to a minimum.

Apparently, therefore, volume diffusion of water from glass is unimportant during life operation of most electron tubes. Consequently, the unusually rigorous exhaust schedules which would be required to deplete water by diffusion are unnecessary. The removal of water from the hydrated surface-layer is important and may be achieved by brief heating of the glass to relatively low temperatures (300 C or higher) during exhaust bake. Weathering of the glass during storage, and the possibility of removing weathering products by hydrofluoric-acid washing, are factors which affect the amount of surface water to be removed during exhaust bake.

Diffusion in Glass

Diffusion of Water in Glass. The diffusion constant of water in glass can be found from the values of m (discussed in Thermal Outgassing of Glass, above) if the initial concentration C_0 is known. For Corning 0080 glass, C_0 has been determined to be 495 cubic centimeters-millimeters mercury pressure per cubic centimeter of glass; thus, a calculation of the diffusion constant for 0080 glass at a temperature of 430 C gives 6.1×10^{-13} cm^2/sec. Similar data are not available for other glasses.

Diffusion of Gases in Glass. Helium permeates glasses readily even at room temperature. The permeation of helium through glass, like that of water, is limited only by its diffusion rate and solubility in glass. These constants have been determined for a number of glasses (reference 1, page 275). Helium permeation rates are generally higher for the high silica glass compositions, such as the low-expansion borosilicate glasses, Vycor and fused silica.

The permeation rates of other gases including hydrogen, deuterium, and neon have been determined at high temperatures. For ordinary applications at lower temperatures, however, the effect of these gases is insignificant.

Diffusion of Ions in Glass. The mobility of the sodium ion in glass has been shown to be responsible for its electrical conductivity at ordinary temperatures. Other ions such as silver, copper, and lithium ions, can replace sodium ions either by diffusion or by electrolysis. This replacement occurs, for example, when a glass surface is in contact with a molten salt containing compounds of these elements.

Some methods for staining glass are based on the use of ion exchange to introduce copper or silver ions into the glass surface. The surface, coated with a clay suspension containing a silver or copper salt, is heated to the annealing range of the glass. During cooling, the silver or copper ions precipitate to form a colloidal dispersion of the metal, and produce a characteristic red or brown color in the surface layer of glass. Brown stains are sometimes produced by the accidental introduction of iron ions during the washing of commercial glass articles.

The surface properties of glasses are altered by ion exchange. Stained glass surfaces may show improved resistance to weathering. In another example, the ability to apply a uniformly conductive coating of tin oxide was lost because the glass surface had been contaminated by migration of iron from a kovar ring to which the glass had previously been sealed.

Surface properties may similarly be altered by diffusion of sodium ions out of the glass. Diffusion can be produced by heating the glass in atmospheres containing acid gases such as carbon dioxide, hydrochloric acid, or sulfur dioxide. These gases react with sodium ions to produce solid soluble sodium salts which subsequently can be washed off the glass surface. A glass surface depleted of sodium ions improved resistance to weathering and may acquire other properties characteristic of the surface of a high-silica glass. Solid salts are sometimes left on the surface after such a treatment when commercial glass articles are annealed; these deposits are referred to as lehr bloom. Sulfur has been used to slightly contaminate the boats used for annealing the target glasses of image orthicons; the contamination produces a light uniform film of sulphate salts on the surface of the targets. Because dust particles apparently do not stick to this surface, scrap due to spots and blemishes in the picture is reduced. Sulfur dioxide is sometimes added in small concentrations to the natural gas used on button stem machines; in this case, the sulfate film formed on the glass surface lubricates the button and helps to prevent the glass from sticking to the molds. Heavy chloride films have occasionally been found on button stems made and annealed in factory areas adjacent to trichloroethylene degreasing units. This undesirable condition should be promptly corrected.

Ion exchange similarly occurs at the seal interface when two glasses are sealed together. If the two glasses are not identical in composition, a transition zone at the interface is produced which usually has properties intermediate between those of the two glasses. This effect ordinarily improves the quality of the seal. In some unusual combinations of glasses, however, the ion exchange shifts properties such as thermal expansion in the wrong direction and the transition zone may contain altered glass layers which are highly strained as a result of interdiffusion during sealing.

Gases Dissolved in Glass

Gases dissolved in glass are determined by heating a sample in vacuum to high temperatures (1200 to 1400 C) and collecting the evolved gases for analysis. (When this method is used, the glass froths and liberates gases rapidly.) The predominant gas found in commercial glasses is water vapor, which is present at less than one cubic centimeter (standard temperature and pressure) per cubic centimeter of glass. The amount of water is believed to be a function of the particle pressure of water in the melting-tank atmosphere, due to an equilibrium which is probably reached during an early stage of the melting cycle. Smaller amounts of carbon dioxide, sulfur dioxide, and oxygen are also found; the quantities of these gases depend on the presence of carbonates, sulfates, and nitrates in the glass batch and also upon the presence of fining agents, such as arsenic oxide, which are added to help remove glass

bubbles from the glass melt. The control of the concentrations of these gases during the melting of glass is important to the glass manufacturer who must eliminate bubbles or seeds in his glass to obtain a product of acceptable quality.

Bubbles and air lines are defects in glass which are introduced or not removed during the melting process. The pressure of gas in the bubble is about one atmosphere at the glass-melting temperature, and changes little until the glass becomes substantially rigid during initial cooling; the "setting point" of the glass is about 900 K. Further cooling of the glass to room temperature reduces the pressure in the bubble to about 1/3 atmosphere. An analysis of gases in bubbles in Corning 9010 glass[16,17] showed that carbon dioxide, nitrogen, or oxygen were the predominant gases; in addition, some bubbles contained argon in amounts of about 1 per cent of the total gas. Some of these bubbles are trapped air; bubbles in soda-lime-silica glasses have been reported to contain substantial quantities of sulfur dioxide.

Outgassing under Bombardment

When glass is bombarded with 20-kilovolt electrons, gases are evolved which are primarily oxygen.[18] The amounts of oxygen found in several glasses are given in Fig. 8. These data were obtained on scanning a 3/4-square-inch raster with a picture tube gun, using a target current of 200 microamperes and a glass surface coated with a thin aluminum film to provide electrical conductivity as a target. The sample temperature was estimated to be about 250 C.

A total quantity of 60 liter-microns of oxygen is liberated from the scanned layer of 0080 glass during a 20-hour test. The 20-kilovolt electrons have a range of about 2 microns in glass. If only this layer of glass at the surface contributes to oxygen evolution, it has been calculated that as much as 10 per cent of the total oxygen in the glass structure was released during the test. The surface layer of glass is altered by bombardment; physical disintegration and roughening of the surface are observed after long operation of some glasses in kinescopes. Discoloration or browning of the surface layer of glass by electron bombardment, however, should not be used as an indication of the amount of oxygen evolution.

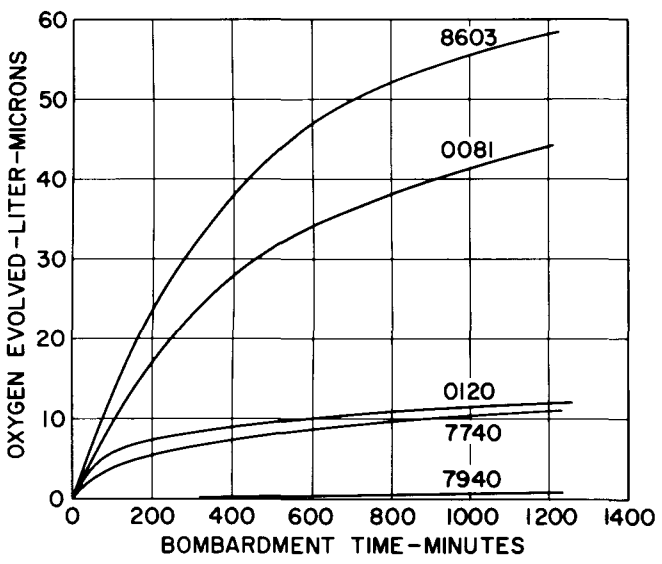


Figure 8. Oxygen Evolution for Five Glasses

The mechanism of oxygen release by glasses under electron bombardment has not been established. Penetration of the glass by electrons results in a charge layer under the surface; the return of this charge to the surface probably is electrolytic in character; if it is electrolytic, then release of oxygen gas at the surface would be expected. Oxygen evolution of a glass, however, appears to be only indirectly related to the electrical conductivity of the glass.

In some cases, glass surfaces which may be subjected to electron bombardment are coated to prevent charging of the surface or other undesirable effects. Most of these coatings do not eliminate oxygen evolution. Thus, the evaporated aluminum film used in cathode-ray tubes is transparent to electrons and, therefore, oxygen released by the glass readily passes through the film. Qualitative data indicate that conductive tin-oxide coatings behave similarly. Coatings bonded with sodium silicate (such as the conductive graphite-silicate coating used in cathode-ray-tube types) have been shown to be prolific sources of oxygen under bombardment.[18]

Data are not available on the variation of oxygen evolution of glasses under electron bombardment for parameters such as glass temperature or voltage of the bombarding electrons. However, some qualitative data taken at RCA-Lancaster showed an exponential dependence on temperature, with the amount of oxygen evolved at 20 kilovolts doubling with a 30 C increase in glass temperature. A substantial evolution of oxygen on low-voltage bombardment (200 to 500 volts) of a graphite-silicate coating has been observed qualitatively.

The practical importance of oxygen evolution in a tube depends upon the conditions of operation. Bulb bombardment, or scanning and back-scattering of electrons, such as occurs in aluminized cathode-ray tubes, is likely to release substantial amounts of oxygen. For these tube types, an efficient getter for oxygen is necessary because the quantities of oxygen evolved considerably exceed the quantity which poisons oxide cathodes. It seems probable that the superior performance of ceramic-envelope, high-frequency power tubes over glass-envelope tubes, may be partly related to less susceptibility to bombardment effects.

Sublimation from Glass Surfaces

Sublimation is observed where glass surfaces are heated to temperatures above 400 C. At the high temperatures used for glass sealing, the evaporated materials contain silica and the other oxides present in the glass. During flame-sealing of glass, a "ring stain" caused by deposition of these oxides on adjacent cool surfaces may sometimes be observed. Once formed, such stains are difficult to remove by chemical cleaning but may be controlled by heating the cool area dur-

ing sealing to prevent the oxides from condensing on it.

During main sealing operations where the final closure is made by means of a glass seal, materials evaporated from the seal may contaminate the oxide cathode or other sensitive tube elements. The glass bulb composition has been reported to have a considerable effect on contamination incurred during sealing and exhaust.[19] When Corning 7052 glass was used for the envelope of a small diode tube severe contamination due to black-oxide coatings and consequent low emission occurred. The use of envelopes of 7040 glass, or a combination of 0080 bulb and 0120 stem, however, resulted in white coatings with good emission and life. The results obtained with 7052 glass were probably due to its fluoride and chloride content. The 7052 contains about 1/2 per cent of fluoride and 1/5 per cent of chloride, while 7040, 0210, and 0080 glasses are essentially free of chloride and fluoride.

In the above case, contamination of the cathode occurred principally during the stem-sealing operation. In other cases, cathode contamination by fluorides evaporating from fluoride-containing bulb glasses has been observed during the exhaust-bake operation.

DEVITRIFICATION OF GLASS

Glass is a noncrystalline metastable form of matter. Commercial glasses generally show a very slow rate of crystallization only in a narrow range of temperatures below the liquidus temperature, which is the temperature at which crystals of the primary phase first appear when a glass melt is slowly cooled. In most commercial glasses, the crystallization range is above the softness temperature; thus, in melting or working these glasses, the crystallization range is passed through so quickly that no crystal formation occurs. Devitrification (crystallization) of these glasses is, therefore, rarely encountered in most practical applications.

Devitrification is a phenomenon sometimes made use of in the manufacture of glass articles. Opal glasses are special compositions in which opalescence is due to the presence of numerous small crystals; these crystals are usually produced in the glass by reheating the article after it has been formed. Photosensitive glasses have been developed by Corning Glass Works in which nucleation centers are produced by exposure of the glass to ultraviolet radiation; on subsequent heat treatment at temperatures in the upper annealing range, a colored image may develop due to the precipitation of colloidal metal particles in the exposed regions. Corning Fotoform glass is a photosensitive glass in which the exposed regions develop into an opal glass upon heat treatment, while unexposed regions remain clear. The opal regions may subsequently be removed by etching in dilute hydrofluoric acid. Due to the selective action of the hydrofluoric acid, holes of minute size or complicated shapes may be accurately reproduced in Fotoform glass from a photographic master. Corning Pyroceram compositions are sensitized glasses which may be manufactured and worked as clear glass and subsequently devitrified by heat treatment in the upper annealing range. Due to the extensive formation of small crystals, these materials have quite different properties from the parent glass, being characterized by high mechanical strength and low electrical loss.

Certain borosilicate glasses may separate on heat treatment into two immiscible glass phases; the manufacture of Vycor (96 per cent silica) glass is based on this phenomenon. Following the heat treating of a suitable borosilicate glass, two glass phases are formed, one being a high silica glass, and the other a soluble borate glass which may be leached away by immersion in an acid. The porous glass structure left after removal of the soluble phase is sintered to a solid glass by further heat treatment, producing a high-silica glass article of similar shape but of smaller size than the original.

Corning 7040 kovar sealing glass is a borosilicate glass and exhibits the same kind of instability on heat treatment. When this glass is held at temperatures somewhat above the annealing point, a similar separation into immiscible phases occurs, one phase being soluble. Thus, if a 7040 glass-to-metal seal is annealed using temperatures substantially above the annealing point, and if the seal is then chemically cleaned or exposed to water, leaching of the soluble phase in the surface layers of glass may occur. This state of the glass cannot readily be determined by visual examination but if the seal is reheated for further sealing or working, a violent re-boil or bubbling of the molten glass will be observed. This condition can be controlled partly by the use of lower annealing temperatures and shorter times, and partly by minimizing exposure to water after annealing. Other borosilicate glasses used for electron tubes generally do not show this behavior.

THERMAL EXPANSION OF GLASS

When glass is cooled from a high temperature, the transition from the liquid to the solid state occurs, not at a fixed temperature, but over a temperature range referred to as the transformation or annealing range. The thermal expansion of glass in both the liquid and solid states depends on the chemical composition; the thermal expansion in the liquid state is generally about three times that in the solid state. Within the transformation range, however, the specific volume of glass is strongly dependent upon its heat treatment; for these temperatures, expansion is determined by the previous history of the glass as well as by its composition.

The thermal-contraction curves typically observed when a glass is cooled at different rates through the transformation range are shown in Fig. 9. These curves, drawn from a common point at maximum temperature, show the same contraction rate for the liquid state. The curves deviate from the "liquid" line in the transformation range with slope decreasing until at the lower limit of the transformation range, the common characteristics of the solid state are attained. The curves for the slower cooling rates follow the "liquid" line to lower temperatures and show greater net contraction in going through the transformation range. Upon further cooling to room temperature, the curves form parallel lines with the slope of the curve for the solid state. Upon reaching room temperature, the glass

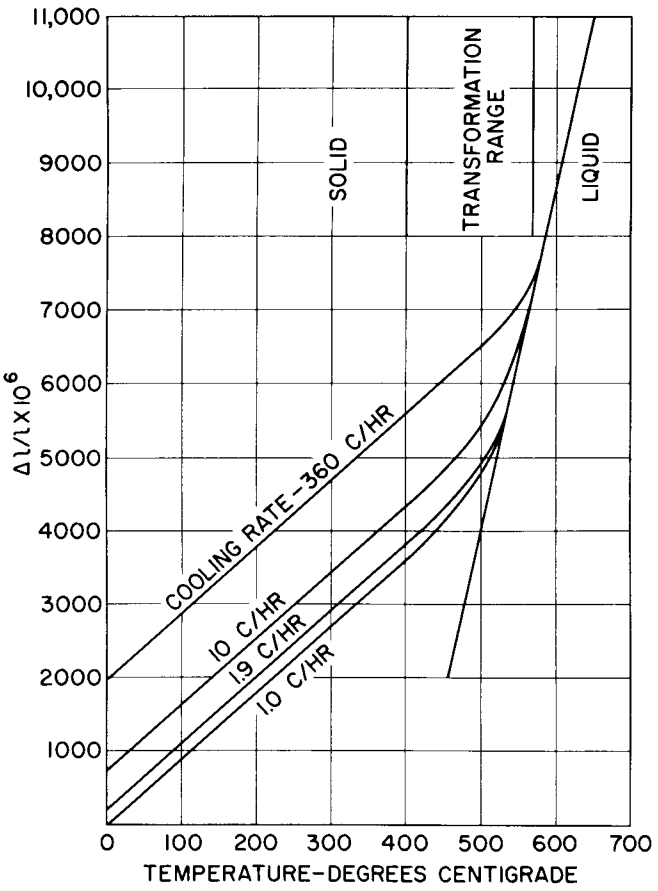


Figure 9. Thermal Contraction Curves of Glass vs. Cooling Rate

samples cooled at the slower rates thus have shown a greater net contraction, as indicated by the displacement of the contraction curves and, therefore, they have lower specific volume and higher density than the rapidly cooled samples. The other physical properties of glass, including density, depend somewhat upon cooling rates of heat treatment used in bringing the glass sample through the transformation range; thus, a defined cooling rate is usually specified in giving the properties of glass. These effects have been discussed with reference to the constitution and structure of glass.[20]

In making glass seals, it is important to match the thermal expansion of the glass to that of the other component of the seal to minimize or control the residual stresses in the finished seal. Thermal expansion must be closely controlled in commercial glasses used in electron tubes. Sometimes, the expansion of a glass has been specified and controlled by the use of a sealing test in which a sample is sealed to a standard material and the residual stresses are determined after a controlled annealing schedule. In most cases, however, the average coefficient of expansion, determined for the temperature range of 0 to 300 C, is used for the specification of this property. This measurement involves only the linear part of the thermal-expansion curve and thus avoids the variable effects which can occur in the transformation range. In practical applications of glass seals, however, the processes occurring in the transformation range are capable of substantially modifying the residual stresses found in the seal. On first annealing the seal the cooling rate has an obvious effect, according to Fig. 9, and further changes in residual stresses may also be found during subsequent heat treatment of the seal in the transformation range;[21] for ordinary glasses, these effects may be seen during heat treatment at temperatures as low as 100 C below the strain point of the glass.

GLASS-TO-METAL AND GLASS-TO-GLASS SEALS

Nature of Sealing Problems

In most applications of glass in electron tubes, the use of glass depends greatly upon its compatibility with the material to which it is to be joined. If the resulting seal is not hermetic through a reasonable temperature range, then it is not acceptable. The properties of the metal also influence the degree to which a seal is dependable. The factors that influence the reliability of a glass-to-glass seal are less complex than those that affect glass-to-metal seals. Usually the thermal expansion of the materials is a major factor for consideration when materials are matched for the construction of components in the electron-tube industry. A close match in expansion is definitely required in order that a seal of dissimilar components may remain intact and be able to withstand the effect of thermal change during use; however, some seals are made with materials deliberately chosen to put the glass member under compression and thereby improve mechanical strength. Other factors (including linearity of expansion, surface tension characteristics, oxidation characteristics of metals, and viscosity characteristics) play an important role in determining whether a given glass may be sealed to another glass or to a specific metal. Other physical and chemical properties, of course, must also be included in the over-all application consideration.

In order that glasses may be sealed to other glasses or to metals, a number of criteria must be satisfied, as applicable:

1. The glass must wet the metal. Any clean glass will wet a clean metal or adhere to it if the metal is covered with an adherent layer of the metal oxide and the temperature of the components is raised to a point where this oxide is partially dissolved in the glass.
2. The stresses resulting from thermal expansion and contraction must not exceed the tensile strength of the glass. The thermal expansion characteristics of the components should be linear and without any points of inflection (transformation points) below the strain point of the glass. It is assumed that the temperature of highest use in the system will be substantially lower than the strain point of the glass.
3. Oxide layers formed either during or prior to the sealing operation must be firmly adherent to the parent metal and, in addition, be of an impervious nature.

In recent years, specific applications have called for the tailoring of the components to suit the needs of the application. The metal kinescope is a case where both

the metal and glass components have been specially developed to meet the requirements of the given application. In this particular application, the metal kinescope was first produced using available materials. The round metal funnel was made by spinning a blank of Type 446 chromium-iron alloy (28 per cent chromium), and the bent faceplate to be sealed to the funnel was made from a selected drawn-sheet window glass. Fabricating the parts for a rectangular type metal tube, however, required a more workable metal and AISI 430 alloy was developed. This metal is a 17 per cent stabilized chromium-iron alloy capable of being formed to the desired shape by both spinning and stamping. Because this alloy had a high thermal expansion, a modified window glass was developed to go with the new metal. This glass (Pittsburgh Plate 3720) had a neutral gray tint to improve picture contrast. Still later, enameling iron was used for the metal funnel and new iron-sealing glasses were developed for the faceplate (as window glass) and for the pressed funnel; machine-drawn tubing was provided for the neck and stem sections.

Adherence of Glass Sealed to Metal. When glass is sealed to metal, an oxide film is generally formed on the metal surface. Glass, itself being a solution of oxides, wets the oxide film during sealing; the mechanism of adherence is commonly regarded as being due to the adherence of the oxide film to the metal surface. The solubility of the oxide film in the glass and the tenacity with which it adheres to the parent metal are thus fundamental to the seal. The color of the seal is determined by the dissolved oxide and is useful in determining the character of the seal. In practice, observation of the depth of the color is effective for determining whether the thickness of the oxide is appropriate. In general, the deeper the color, the thicker the layer of oxide and it is possible, therefore, to establish whether the oxide thickness has exceeded or fallen short of prescribed limits. The oxide thickness may be controlled by appropriate atmospheres during sealing if over-oxidation is a problem or by means of preoxidizing if the metal oxidizes with difficulty. Cold-rolled steel applications require oxidation protection during the sealing operation because of their affinity for oxygen. Conversely, chrome-bearing alloys must be pre-oxidized under specific conditions so that an optimum oxide layer may be formed. Partridge[22] offers detailed information on the composition and properties of oxide films and their effect upon glass-to-metal seals.

Estimating Quality of Seals. The quality required of a given seal depends upon the specific application and the conditions to which a component is subjected in use. However, some general considerations can be cited and will be found to be applicable in most situations. Figs. 10 and 11 illustrate criteria that were applied successfully to the examination of metal kinescope bulbs for such types as the 21AP4 and its predecessors. Foremost is the wetting angle or angle of approach of the glass and metal. If the re-entrant angle is less than about 50 degrees, the strength of the seal will be adversely affected. A small contact angle is indicative of good wetting characteristics and is usually desirable.

Colorimetric determination of seal quality is frequently used, as mentioned previously. The seal color directly indicates the quality of the oxide layer. Thus, a light-colored or metallic seal may show poor adherence of glass to metal, while a dark-colored seal may leak due to the porosity of an excessively thick oxide layer.

Polariscopic examination of the seal affords an indication of the nature of the stresses in the seal and indicates whether the system of materials being employed is proper or if the process is under control.

Glassy seals should be avoided because of the tendency of such seal areas to coalesce and form paths for both electrical and gaseous leakage. In addition, such defects often lead directly to mechanical failure in the seal. Gassy seals are characteristic of certain metals, notable among which is nickel.* Where design criteria indicate the desirability of such a material, it is best to provide for plating or otherwise coating the metal so that the coating material predominates the seal character. Another means for avoiding any tendency toward gassy seals is to outgas the metal (prior to making the seal) at a temperature at least 50 C higher than the maximum temperature reached during the making of the seal.

Blister, scuffs, and opaque spots should be critically examined in the evaluation of seal quality. Blisters usually result from the elongation of a gaseous inclusion in the glass. Scuffs are surface blemishes made up of a series of parallel scratches that are so fine that the unaided eye cannot separate them. Opaque spots are usually attributed to either included material in the glass or else to oxide scale that has worked free from the metal and found its way into the seal. The presence of any of these three types of defects is not immediate cause for rejection of the seal in itself, but it should be recognized as a possible contribution to seal failure when setting up defect limits for specific seals that can be used successfully. Rather than cite specific limits, it is suggested that the limits be determined for each individual case.

Seal geometry plays a very important role in the evaluation of the quality of a seal. Obvious cases of re-entrancy between the metal and the glass should be avoided because of the tendency for stresses to concentrate at such points of discontinuity. Geometry requirements include the need for seal uniformity; a one-sided seal, for example, will generally invite failure.

Surface defects should be avoided primarily because of the fact that where tensile forces are present they tend to concentrate at surfaces and will inadvertently seek out these surface defects with consequent seal failure.

Checks or very small cracks at or near a glass-to-metal interface should be cause for immediate rejec-

* Gassy materials are deliberately used for leads or studs which are anchored in a glass piece but which are not sealed to the glass, e.g., the nickel parts on stem leads.

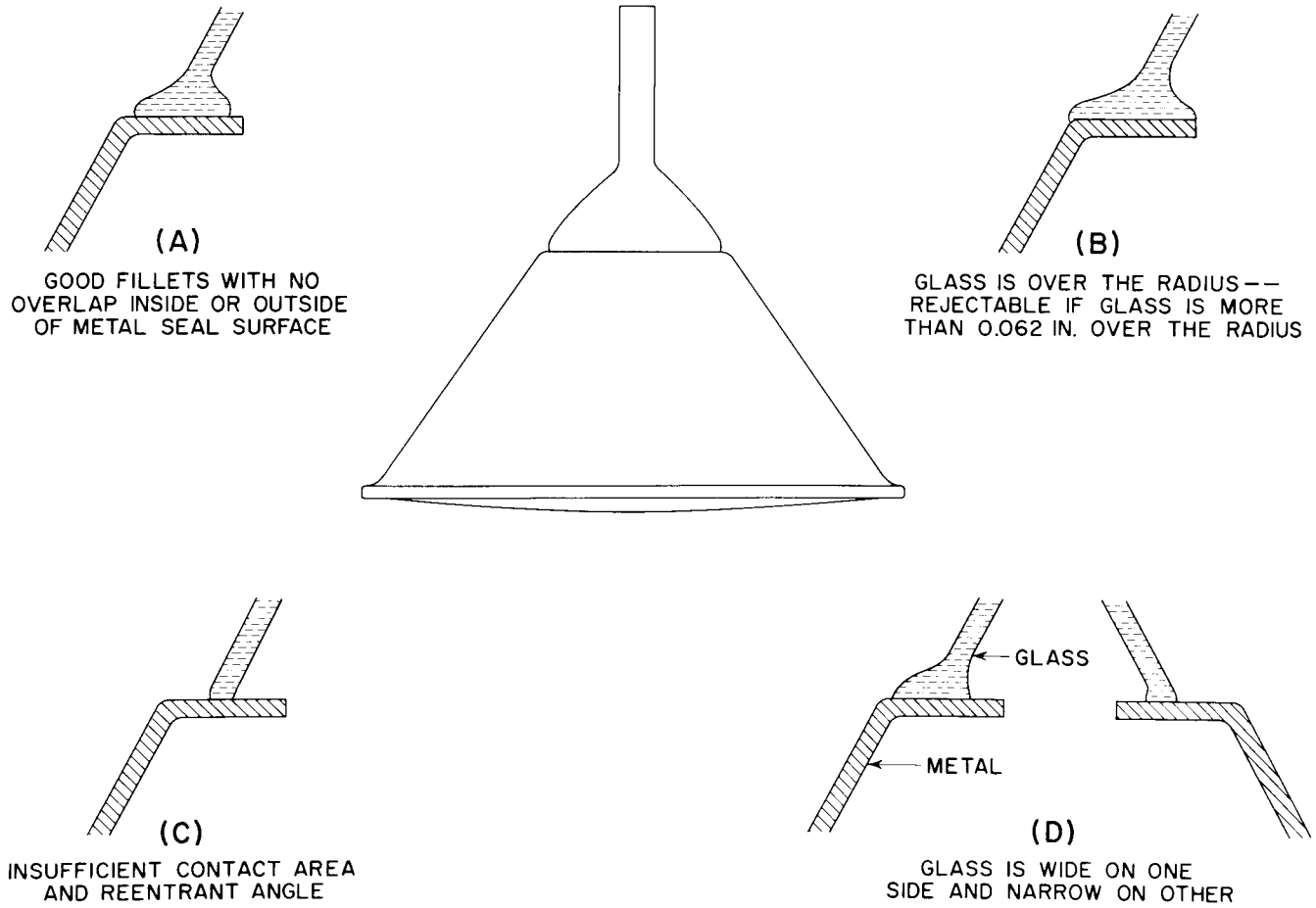

(A)
GOOD FILLETS WITH NO OVERLAP INSIDE OR OUTSIDE OF METAL SEAL SURFACE
(B)
GLASS IS OVER THE RADIUS -- REJECTABLE IF GLASS IS MORE THAN 0.062 IN. OVER THE RADIUS
(C)
INSUFFICIENT CONTACT AREA AND REENTRANT ANGLE
GLASS
METAL
(D)
GLASS IS WIDE ON ONE SIDE AND NARROW ON OTHER

Figure 10. Examples of Seal Geometry—Metal Kinescope Funnel Seals: (A) good seal, (B) wide seal, (C) narrow seal, (D) uneven seal

tion of a seal. The defect may not have resulted from the stresses that occur as a result of the mismatch between the components, but it is inevitable that further failure from this mismatch will originate from the observed check or crack.

Commonly Used Glass-to-Metal Seal Combinations. A general summary of suitable combinations of glasses and metals which have proved to be successful in application to electron tubes has been compiled. Table IV shows a list of the combinations discussed here.

Dumet is the most widely used wire-lead material in the electron-tube industry. It is essentially a 42 per cent nickel-iron alloy coated with a sheath of copper which occupies 20 to 25 per cent of the cross-sectional area of the wire. It is sealed in combination with Corning 0080, 0120, 0010, 8160, and 8161 glasses and other glasses of similar expansion characteristics. Dumet seals are made in diameters up to 0.040 inch but the more common applications are in diameters below 0.025 inch. The seals are usually free of any gaseous inclusions and the red copper oxide is easily dissolved in the sealing glasses. The seals can be made in air atmospheres with relative simplicity. It is generally found that a gradation in seal color develops, ranging from the metallic luster of pure copper to that of the deep red oxide. Gradation within this range will insure that an optimum leak-free seal condition will exist.

52 Alloy contains approximately 50 per cent nickel and finds some application in tube construction in seals to the glasses described above under dumet wire.

Chromium-Iron Alloys have received wide application in recent years; their technology is becoming extensive; these alloys are essentially stainless steels containing either 17 or 28 per cent chromium. They offer a relatively low-cost source of oxidation-resistant material that can be sealed easily to soft glasses. The incidence of allotropic transformations in these alloys was initially a problem, particularly with decreasing chromium content. Suppression of this tendency toward transformation was made possible by the addition of varying amounts of aluminum, molybdenum, titanium, vanadium, tungsten, or tantalum. The alloy known in the industry as 430TI (AISI Type 430, modified) is an example of a composition that would ordinarily exhibit a tendency toward austenitic transformation and which was stabilized through the addition of titanium. Both 446 and 430 alloys can be sealed to 0120 glass and glasses of similar expansion characteristics. However, the stresses resulting from thermal expansion differential are fairly high; for large or critical seals, special

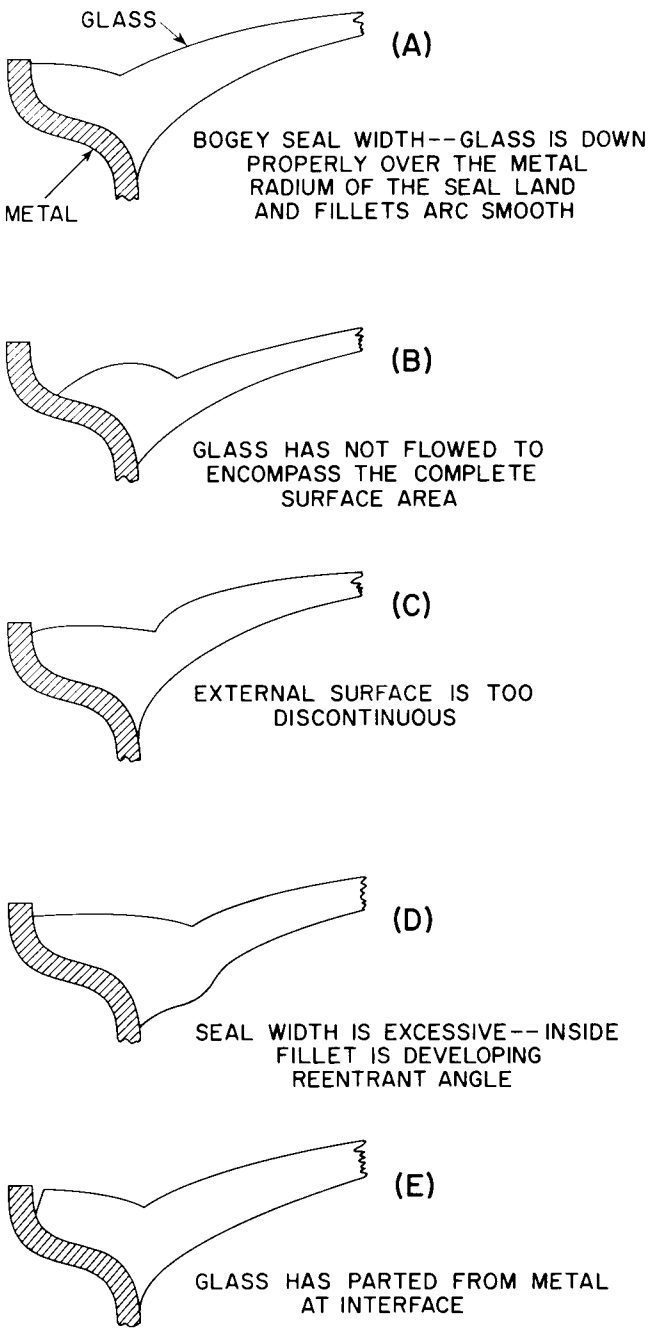

GLASS
(A)
BOGEY SEAL WIDTH--GLASS IS DOWN PROPERLY OVER THE METAL RADIUM OF THE SEAL LAND AND FILLETS ARC SMOOTH
METAL
(B)
GLASS HAS NOT FLOWED TO ENCOMPASS THE COMPLETE SURFACE AREA
(C)
EXTERNAL SURFACE IS TOO DISCONTINUOUS
(D)
SEAL WIDTH IS EXCESSIVE--INSIDE FILLET IS DEVELOPING REENTRANT ANGLE
(E)
GLASS HAS PARTED FROM METAL AT INTERFACE

Figure 11. Examples of Seal Geometry—Metal Kinescope Faceplate Seals: (A) good seal, (B) cold seal, (C) sharp seal, (D) wide seal, (E) stripped seal

glasses of higher thermal expansion have been developed. Among these are Pittsburgh Plate 3720 and 5533 faceplate glasses, Lancaster Glass G12X1, and Corning 9019, 0129, and 0128 glasses.

Sylvania No. 4 Alloy or Sealmet 4 is a nickel-chromium-iron alloy used extensively in electron-tube construction. It contains 42 per cent nickel and 6 per cent chromium and exhibits an expansion curve having a close coincidence with Corning 0120 glass. The alloy forms a very adherent oxide when it is heated to 1100 C in a wet hydrogen atmosphere. The chromium-oxide layer is readily soluble in glass and exhibits a characteristic green color.

Nickel has found little application in seals in electron tubes because of its high thermal expansion. It is used extensively in conjunction with dumet in the form of a two- or three-piece lead assembly. In this role, the electrical conductivity of the material is employed, as well as its ductility. Nickel is readily manipulated to form supports for tube components. In this application, the dumet section of the lead wire is sealed into the glass and forms the hermetic seal; the nickel sections are buried in the glass and thus, are held mechanically although there is no adherent seal to the glass.

Copper, because of its high thermal expansion, cannot be matched to the common glasses. However, with a Housekeeper, or feather-edge seal, it may be sealed to any glass. Recently a glass that matches copper has been developed and is available in the form of rod, powder, or sintered beads; this glass is called Corning 7295 copper-sealing glass and may find application to electron tubes.

Kovar (also known as Rodar and Therlo) is a nickel-cobalt-iron alloy developed for sealing to glasses of lower thermal expansion.[23] Glasses made for sealing to kovar are Corning 7040, 7052, and 7056, and Kimble L-650 glass. Kovar seals well to these glasses; it has an expansion curve which parallels the expansion curves of these glasses over the range of temperatures encountered in sealing and in use.

Tungsten can be sealed to a wide variety of hard glasses including Corning 7720, 7780, 3320, 7070, 7750, and 7760. Tungsten, like molybdenum, is a pure, hard, brittle metal, capable of being sealed directly to these glasses. Its brittleness requires careful working of the metal to avoid splitting or breaking it. Tungsten oxidizes readily at comparatively low temperatures; a thin coating adheres well. This material finds wide application for lead seals in power tubes. Use of this material in tube construction requires caution because of the chemical instability of its oxide in the presence of water vapor, as encountered in the operation or storage of tubes under conditions of high humidity. This instability sometimes leads to slow-leakage conditions which are not detectable under ordinary circumstances but which will show up in use. Inspection of drawn tungsten wire must be meticulous because, in the drawing process, defects in the material may become elongated into fissures and produce leaks in finished seals. In spite of these difficulties, tungsten serves the purpose and is used widely in many applications in the electron-tube industry.

Molybdenum can be matched to kovar-sealing glasses and hard glasses of similar thermal expansion characteristics, such as Corning 1720 glass. The affinity of molybdenum for oxygen poses a problem; in fact, the oxidation problem is so great that the application of molybdenum to electron tubes has been limited until recently. Molybdenum begins to oxidize at 200 C, and at 400 C it readily forms molybdenum trioxide; this oxide becomes volatile at about 800 C. In small seals

Table IV

Combinations of Glasses and Metals Used For Sealing

Sealing Metal	Sealing Glasses		Expansion Coefficient (cm/cm/°C)
Tungsten	7761	(Corning) Multiform	28×10^{-7}
	7070	(Corning)	32
	7740	(Corning)	32.5
	7760	(Corning)	34
	7720	(Corning)	36
	K-772	(Kimble)	36
	7780	(Corning)	37
	9741	(Corning)	39
	3320	(Corning)	40
	7750	(Corning)	40.5
	1720	(Corning)	42
Kovar, Molybdenum	7050	(Corning)	46
	7052	(Corning)	46
	K-705	(Kimble)	47
	7040	(Corning)	47.5
	K-704	(Kimble)	49
	K-650	(Kimble)	51
	7056	(Corning)	51
	BSC-51	(Bausch and Lomb)	51
	7055	(Corning)	51.5
	Most Commercial Window and Container (Bottle) Glasses		80 to 85
Dumet, No. 4 Alloy, 52 Metal	9010	(Corning)	88.5
	0120	(Corning)	89
	KG-12	(Kimble)	89
	G-12	(G. E., Lancaster)	89
	FR-60	(Fourco)	90
	K-30	(Kimble)	90
	3459	(Pittsburgh)	90
	0010	(Corning)	91
	8160	(Corning)	91
	HR	(G. E.)	91
	G12X1	(Lancaster)	91
	0080	(Corning)	92
	0083	(Corning)	92.5
	R-6	(Kimble)	93
446 Alloy	G-8	(G. E.)	95.5
	5533	(Pittsburgh)	96
	3720	(Pittsburgh)	96.5
430 Alloy	0129	(Corning)	98.5
	9019	(Corning)	99
Iron	LL117	(Lancaster)	118
	IS115	(G. E.)	118
	4442	(Pittsburgh)	118
	7210	(Pittsburgh)	119
	R-3	(G. E.)	124
	1991	(Corning)	128

(such as wire seals) this characteristic has not been detrimental because the oxide formation could be either avoided or held below detrimental limits; however, large seals had to be avoided. The use of nitrogen atmospheres in the making of molybdenum seals proved unsuitable because the metal formed a brittle nitride. Until recently, the primary solution to this problem was to prebead the molybdenum part in an inert atmosphere (argon) and to make the final seal under conditions that would not cause further oxidation of the metal.

This approach eliminated the possibility of larger seals because of the inconvenience in prebeading a large part. More recently, methods have been developed in which the molybdenum surface is either converted to molybdenum disilicide in one process or is enriched with chrome in another process.[24] Neither of these surfaces possess the affinity for oxygen that molybdenum does and permit seals to be made directly to them.

Low-Carbon Steel presents considerable difficulty in sealing because of the ease with which the metal oxidizes. Protection of the metal surface during exposure to high sealing temperatures is a definite requirement and can be accomplished by plating the surfaces with copper, silver, chromium, or nickel. In metal-kinescope production, the seal surfaces were protected during the glass-sealing operation by the use of a vitreous enamel. This coating was applied either by spray or dip application and then was fired in an atmosphere at a temperature of 800 to 900 C. The use of this enamel assisted the glass to flow and form fillets more readily than otherwise would have been the case. Because of the high thermal expansion of these metals, only special glasses can ordinarily be sealed to them. Corning 1991, General Electric R3 and IS115 glasses, Lancaster Glass LL117, and Pittsburgh Plate 4442 and 7210 glasses have been developed for sealing to iron.

Sealing Methods. The art of glass sealing has progressed greatly from the simple process where two pieces of glass of the same composition were joined by heating them in a flame. In today's technology, seals can be made in almost any form as dictated by the needs of the process under consideration. Methods have progressed from the simple use of the hand torch and bunsen burner to the use of huge lehrs and induction furnaces to make a seal between glasses or between a glass and a metal. Although flame sealing is still the most widely used method, a transition to electrical heating or lehr heating is in process; eventually, these methods will displace flame sealing as the prime means for sealing glasses.

In flame sealing, the object is to bring the two glasses to a temperature at which they both have a similar viscosity. When this condition is attained, the components are joined by bringing them together. If one of the components to be joined is a metal, the metal must be brought to the temperature where the glass flows out over the metal surface and adequately wets the metal. The type of metal used dictates whether the metal is exposed to the flame or whether it must be shielded from the flame, depending on the nature of the metal and its oxide. If the metal oxidizes easily, then it is necessary to prevent the buildup of a heavy oxide layer by covering the metal with glass at the earliest possible moment. On the other hand, if the metal resists oxidation then it may be found necessary to expose the metal to the flame for a prolonged period or even preoxidize it before covering the metal with the glass.

The conditions for making a seal when induction heating is used are similar to those used in flame sealing, particularly with respect to the oxidation problem. In induction heating, radio-frequency energy is applied by suitable coupling to heat the metal component by eddy currents induced in the metal. The heat developed in the metal is transferred to the glass and fusion takes place. This process is used extensively in power-tube production.

Electric-arc sealing, which is not as widely used by tube manufacturers as induction heating, is used chiefly in the sealing of glass bulbs for kinescope application. In this process, the glass surface to be sealed is preheated by a flame. When this surface becomes electrically conductive, a high voltage is impressed through the applied flame to heat the surface by current conduction. With this method, the heat applied is extremely localized and has less tendency to cause distortion in adjacent areas.

Frit, or solder-glass sealing, has come into greater prominence during recent years through the development of a great variety of materials for this use. Among recent applications is the production of the 21CYP22 color kinescope. A high-lead frit-slurry is extruded on the seal edges of the funnel of the tube and the panel is then precisely positioned in contact with the frit. The assembled unit is passed through a lehr and brought up to 445 C for one hour. The frit melts, flows over the adjacent components and then devitrifies into a crystalline solid. The resulting seal is almost as strong mechanically as the bulb; after the seal is made, further thermal processing of the bulb is possible. The seal is impervious to gases and is not electrically conductive. Other developmental frits are conductive but, as yet, these materials are only available for application to the high-expansion glasses; however, applications to low-expansion materials should be forthcoming in the near future.

In general, seals can be classified as being matched or unmatched. In the unmatched variety, they will generally conform to the compressive type in which the glass is the member that is under compression. Compression seals have been used in the construction of metal kinescopes where the linear expansion differential between the glass (measured over the temperature range 0 to 300 C) and the metal was as much as 2×10^{-6} per deg C. Upon cooling, the contracting metal induced a high degree of compression into the glass, the effect of which was to allow rapid heating and cooling rates during tube processing. A similar situation is found in a semiconductor stem for which a dumet or nickel-iron lead wire is sealed into a glass bead which is simultaneously sealed into a small cold-rolled steel header. The application of high compression permits the use of materials which ordinarily are not compatible.

The housekeeper seal is one in which unmatched materials are sealed together using a metal member that is feathered down to a very thin edge; with proper ductility of the metal, the stresses that result from expansion differences are dissipated through the thin section of the metal.

Graded seals are those made between glass components whose difference in expansion characteristics requires a gradual transition from one expansion range to another. Usually, the gradations are made with intermediate glasses, although they are sometimes made

by brazing or welding glass-sealing metals of different expansion characteristics together. Where possible, the glasses in a graded seal are stepped in thermal-expansion-coefficient units of about 0.5 to 1.0 x 10^{-6} per deg C.

Stresses in Seals

Calculation of Stresses from Thermal-Expansion Differentials. Estimating stresses from thermal expansion data for the materials of a glass-to-metal or a glass-to-glass seal is usually reliable and serves as an approximation of the values to be expected in applications. The stress left in a seal is determined by the difference between the thermal-contraction curves of the seal materials, and is taken from the annealing point of the glass or glasses to room temperature. For calculation of residual stresses, a setting point concept is used; this is defined as the point in the annealing range at which these contraction curves must be coincident to explain the observed residual stress. The setting point can be regarded as a fictitious freezing point of the glass.

Tables V and VI give expansion differentials for various combinations of metals and Corning glasses in common use in sealing applications. Directions for their use and an explanation of the units and symbols, as supplied by Corning Glass, are listed below.

1. Differentials in the expansion curves are tabulated in parts per million at the setting point of the glass as in the case of a glass-to-metal seal, or at the setting point of the softer glass as in the case of a glass-to-glass seal.
2. The setting point has been arbitrarily defined as the temperature midway between the annealing and strain points.
3. A number with a negative sign directly above it indicates that the material listed in that row has a lower effective expansion than the material listed in that column.
4. Values in excess of 1000 parts per million are indicated by a positive or negative sign.
5. Metal-to-metal combinations are of no interest and are indicated by a large X.

"Caution should be observed in any attempt to predict residual stress in a seal from the differentials given in these tables, because for any particular seal the setting point is a function of:

Table V

Expansion Differentials - Low Expansion Range*
(microinches per inch)

Material	7070	7740	7760	7720	Tung.	7780	3320	7750	1720	7052	7050	Moly.	7040	Kovar	7055
7070 Glass	0	400	245	55	-175	-35	50	-565	-195	-770	-480	-585	-700	-760	-645
7740 Glass	-400	0	-230	-555	-505	-475	-460	-780	-560	----	-880	-985	----	----	----
7760 Glass	-245	230	0	-305	-345	-305	-185	-670	-375	-895	-725	-780	-890	----	-995
7720 Glass	-55	555	305	0	-20	-105	135	-540	-50	-760	-540	-460	-730	-865	-715
Tungsten	175	505	345	20	0	105	105	-210	-440	-440	-300	X	-455	X	-630
7780 Glass	35	475	305	105	-105	0	115	-500	-125	-715	-455	-525	-660	-720	-595
3320 Glass	-50	460	185	-135	-105	-115	0	-445	-140	-675	-525	-555	-685	----	-820
7750 Glass	565	780	670	540	210	500	445	0	-205	-200	40	-175	-130	-200	-155
1720 Glass	195	560	375	50	440	125	140	205	0	-430	-285	-210	-440	----	-605
7052 Glass	770	+	895	760	440	715	675	200	430	0	260	50	80	5	65
7050 Glass	480	880	725	540	300	455	525	-40	285	-260	0	-110	-205	-275	-160
Molybdenum	585	985	780	460	X	525	555	175	210	-50	110	0	-50	X	-200
7040 Glass	700	+	890	730	455	660	685	130	440	-80	205	50	0	60	30
Kovar	760	+	+	865	X	720	+	200	+	-5	275	X	-60	0	140
7055 Glass	645	+	995	715	630	595	820	155	605	-65	160	200	-30	-140	0

*From Corning Glass Works

Table VI

Expansion Differentials - High Expansion Range*
(microinches per inch)

Material	Plat.	446	9010	42-6	0120	8160	0010	0080	430	Iron	1990
Platinum	0		-420		-240	-360	-320	-510			--
446 Alloy		0	185		340	225	255	215			--
9010 Glass	420	-185	0	65	70	- 25	- 20	270	-340	--	--
42-6 Alloy			- 65	0	40	- 60	- 50	280			--
0120 Glass	240	-340	- 70	- 40	0	- 90	-100	115	-500	--	--
8160 Glass	360	-225	25	60	90	0	- 10	230	-380	--	--
0010 Glass	320	-255	20	50	100	10	0	195	-410	--	--
0080 Glass	510	-215	-270	-280	-115	-230	-195	0	-430	--	--
430 Alloy			340		500	380	410	430	0		-950
Iron			+		+	+	+	+		0	-140
1990 Glass	+	+	+	+	+	+	+	+	950	140	0

*From Corning Glass Works

A. The geometry of the seal.
B. The cooling rate used in the annealing of the seal. The cooling rate used in the determination of the expansion curves was one centigrade degree per minute.
C. The difference in the annealing ranges as in the case of a glass-to-glass seal. A seal made between glasses having similar annealing ranges will have, in general, a lower setting point than that assumed in the tables.

Despite these reservations, these tables should be useful guides in making decisions in sealing problems. As a general rule, any differential less than 100 parts per million can be interpreted to mean a very good sealing condition. Good seals can be made between materials with differentials ranging from 100 to 500 parts per million, but if two materials, having a differential at the higher end of this range, are chosen for a sealing application, the risk of failure is of course greater and, therefore, more care should be exercised in making the seal. Materials with differentials from 500 to 1000 parts per million can be used for progressively smaller and thinner seals.

Measurement of Stresses in Seals. The measurement of stresses in seals in which glass is at least one of the primary members is usually carried out with the aid of a polariscope. The amount of retardation observed in the seal is directly associated with the stress in the seal. By use of the polariscope, a quantitative measurement can be made in many cases. In some instances, the resulting strain pattern is so complex that the observations can be used only as a relative means for controlling an operation. Where the pattern is not complex, the angular rotation required to cause extinction is noted and applied in the calculation of the observed stress, as discussed in the section on polariscopic measurement of stress in glass.

Annealing of Seals

Practical Annealing Schedules. The purpose in annealing a glass seal is to achieve a condition in which the least possible stress remains in the glass. In the annealing process, the glass is heated into its annealing range and held there until stresses developed in the sealing operation have been relieved. In most tube applications, the annealing process follows immediately after the sealing operation before the material has been allowed to cool. This method is more economical because frequently the seal will not stand up without annealing. Once the stresses have been relieved, a cooling schedule is chosen that prevents the development of additional stresses during cooling. In some applications, the glass temperature is raised above the annealing temperature to relieve the stresses more rapidly. The cooling rate depends on glass thickness and the nature of the mismatch of expansion rates between the components in the seal. In the case of thick glass, it might be desirable to cool as slowly as 3 degrees per minute when coming down from the strain point of the glass. However, in most electron-tube applications, the thickness of the glass is less than one-quarter inch and the glass will tolerate more rapid cooling. Rapid rates of cooling can easily be tolerated in thin glass pieces having a symmetrical shape.

Rapid Annealing Schedules for Partial Tempering of Seals. The annealing of the large seals in metal kine-

scopes is done by putting the sealed part directly into an annealing oven having a temperature 20 C above the annealing point of the glass. After a 20-minute stabilizing at this temperature, the envelopes are transferred directly into an atmosphere having a temperature of 25 C. These envelopes are not annealed completely but the faceplates are tempered. The center tension values present in the faceplates of these kinescopes have a bogey value of 120 millimicrons per inch, corresponding to surface compression of about 280 psi. The tempering action permits a more rapid cooling rate to be used than if the glass is completely annealed.

REFERENCES

1. Shand, E. B., Glass Engineering Handbook, McGraw-Hill, New York, 1958
2. Charles, R. J., "Static Fatigue of Glass," Jour. Appl. Phys., Vol. 29, p. 1554, 1958
3. Baker, T. C., and F. W. Preston, "Fatigue of Glass Under Static Loads," Jour. Appl. Phys., Vol. 17, p. 170, 1946
4. Shand, E. B., "Stress Behavior of Brittle Materials," Bull. Am. Cer. Soc., Vol. 38, p. 653, 1959
5. Oughton, C. D., "Analysis of Glass Fractures," Glass Ind., Vol. 26, p. 72, 1945
6. Timoshenko, S., Theory of Plates and Shells, McGraw-Hill, New York, 1940
7. Morey, G. W., Properties of Glass, Reinhold, 1954
8. Goranson, R. W., and L. H. Adams, "A Method for the Precise Measurement of Optical Path Differences, Especially in Stressed Glass," Jour. Franklin Inst., Vol. 216, p. 475, 1933
9. Poritsky, H., "Analysis of Thermal Stresses in Sealed Cylinders and the Effect of Viscous Flow During Anneal," Physics, Vol. 5, p. 406, 1934
10. Sutton, P. M., "Stress Measurement in Circular Cylinders," Jour. Amer. Cer. Soc., Vol. 41, p. 103, 1958
11. Law, R. R., "Contrast in Kinescopes," Proc. IRE, Vol. 27, p. 511, 1939
12. Kreidl, N. J., and J. R. Hensler, "Formation of Color Centers in Glasses Exposed to Gamma Radiation," Jour. Amer. Cer. Soc., Vol. 38, p. 423, 1955
13. Sun, K. H., and N. J. Kreidl, "Coloration of Glass by Radiation," Glass Ind., Vol. 33, pp. 511, 589, and 651, 1952
14. Todd, B. J., "Outgassing of Glass," Jour. Appl. Phys., Vol. 26, p. 1238, 1955
15. Todd, B. J., "Equilibrium Between Glass and Water Vapor at Bake-Out Temperatures," Jour. Appl. Phys., Vol. 27, p. 1209, 1956
16. Todd, B. J., "Mass Spectrometer Analysis of Gases in Blisters in Glass," Jour. Soc. Glass Tech., Vol. 40, p. 32T, 1956
17. Dalton, R. H., "Gases in Glass," Jour. Amer. Cer. Soc., Vol. 16, p. 425, 1933
18. Todd, B. J., J. L. Lineweaver, and J. Kerr, "Outgassing Caused by Electron Bombardment of Glass," Jour. Appl. Phys., Vol. 31, p. 51, 1960
19. Kern, H. E., and E. T. Graney, "Thermionic Emission from Oxide Cathodes as Related to Glass Envelope Composition," ASTM, Special Publication No. 246, p. 167, 1958
20. Condon, E. U., "Physics of the Glassy State. II. The Transformation Range," Amer. Jour. Phys., Vol. 22, pp. 132-142, 1954
21. Turnbull, J. C., "Tension Stresses in Glass Coatings and in Glass-Metal Seals in the Annealing Range," Jour. Amer. Cer. Soc., Vol. 41, p. 372, 1958
22. Partridge, J. H., Glass-to-Metal Seals, Society of Glass Technology, 1949
23. Scott, H., "Recent Developments in Metals Sealing into Glass," Jour. Franklin Inst., Vol. 220, p. 733, 1935
24. Rawson, H., "A Note on Oxidation Resistant Coatings on Molybdenum and Their Use in Glass Seals for Mercury Lamps," Jour. Glass Tech., Vol. 39, p. 211T, 1955

Ceramics

W. F. Lawrence

Harrison

In most dictionaries, even fairly recent ones, "Ceramics" is defined as, "those articles of manufacture made from raw, earthy materials which obtain their durability and usefulness as a result of having been heat-treated at high temperature." Because the field has grown so rapidly in recent years, this description is obsolete. Many of the new ceramic materials have been synthesized by man and are unknown in nature. Such things as silicon carbide, the nitrides, and borides do not even fit the more modern concept of ceramics as being the science of the silicates and oxides. It may, therefore, be easier to approach a working definition from a reverse viewpoint; thus we may say: "Ceramics are those refractory materials and articles which are nonorganic and nonmetallic in character; they are usually polycrystalline, but glass (a noncrystalline supercooled liquid) and certain single-crystal materials such as sapphire are also considered in the field; in almost every case, temperatures above 1000 C are required for their processing and manufacture."

There are many sources (see References) of well-organized data on ceramic materials and their properties, but there are few references, if any, which provide the tube designer with information on how to interpret and apply these data in a practical and meaningful way. This chapter will provide some of this background.

WHEN TO USE CERAMICS

The designer of receiving tubes works primarily with two kinds of materials: Those which are electrically conducting and those which are not. The metal conductors form the "active" part of the tube, whereas nonconducting mica and glass are used as structural members to isolate, support, and contain the metal elements.

As insulators, mica and glass have had an impressive record. They have made the modern electron-tube industry possible. They are low in cost and have many desirable properties. Extensive manufacturing facilities are available to make tubes with these materials. For these reasons the designer must have good justification for recommending new types of insulation or new construction techniques.

Today, severe demands are being made on the electron tube. The limitations of the more traditional materials and methods are becoming apparent. Power and transmitting tubes, for example, in the period since World War II have gone from mica to machined lava and, finally, to steatite or alumina spacers.

Now, with the advent of successful metallizing which makes possible vacuum-tight ceramic-to-metal seals, many power and transmitting tubes are being redesigned to use ceramic spacers, ceramic stems, and ceramic envelopes. The resulting tubes are better than their prototypes in that they can be baked out and operated at higher temperatures; they are smaller and more rugged; they have higher dissipations and better high-frequency characteristics.

The receiving-tube designer is only beginning to make use of ceramics. Present trends seem to indicate, however, that the ceramic receiving tube will be in wide use in a number of markets within a relatively few years.

On what basis, then, can the designer decide to abandon the old in favor of the new? The following check list may be of help. On the basis of comparison with what can be obtained with mica and glass, does the design require:

1. Higher operating or ambient temperatures?
2. Higher bakeout and degassing temperatures?
3. Operation at higher frequency?
4. Higher mechanical strength and ruggedness?
5. Closer spacings made possible by new assembly techniques using high-temperature brazing?
6. Freedom from mica and mica spray particles?
7. Better electrical properties at elevated temperatures?
8. Greater resistance to electrolysis and high-voltage breakdown?
9. Increased uniformity from purer, more reproducible, man-made materials?
10. Materials of lower gas content and vapor pressure?
11. Improved resistance to radiation damage?
12. High temperature printed circuits?
13. New or controlled modifications of thermal conductivity, thermal expansion, surface texture, porosity?

When enough of the preceding questions can be answered affirmatively, the designer is justified in considering ceramics.

It is only in rare cases, however, that a direct substitution can be made in an existing design. Usually, both the design and assembly techniques must be modified substantially. When a ceramic spacer is substituted for mica in a conventional tube type, for example, the advantage of mica flexibility is lost. Aids such as spring clips must be provided to keep the ceramic spacer centered in the bulb. The mount structure must be modified so that the elements are held independently because the ceramic does not have the gripping effect of a feathered mica hole. Balanced against increased cost and complexity, however, are the possibility of higher-frequency operation and greater freedom from particles. Each case is different and must be analyzed very carefully.

It is usually more satisfactory to think of ceramics in terms of new designs or concepts rather than to think of ceramics as a direct substitute for some other material. Until recently, the advantages of ceramics have been nullified by the higher cost of parts and assembly. Simplified, more reliable techniques for metallizing, and the availability of precision ceramics at lower cost give promise of achieving a competitive product, especially when overall assembly advantages and improved performance are considered.

To use ceramics competitively, the designer must make compromises in his choice of material and in the final shape of the part. No one material combines all of the desired properties, and the desirability of the shape of the part usually must be balanced against the cost of fabricating it. The wisest course is to work closely with the ceramic specialist from the beginning so that the requirements can be most efficiently achieved.

THE INTERRELATIONSHIP OF FORMING METHODS AND PART DESIGN

The practicality and cost of any part is a function of how well it can be adapted to a specific forming method, the dimensional tolerances necessary, and the number of pieces required.

Basically, four different methods may be used to obtain a finished, usable ceramic piece. These methods are:

1. Form the piece by one of several standard ceramic techniques to the required geometry (allowing for dimensional changes which occur during drying and firing) and finish-fire to the final size and shape.
2. Form a blank or partially shaped piece. Machine it before firing (allowing for dimensional changes). Fire to final size and shape.
3. Form a blank or partially shaped piece. Partially fire it. Machine it (allowing for dimensional changes). Fire to final size and shape.
4. Form a blank or partially shaped piece. Fire it. Machine it to final size and shape.

Method 1, which is in a sense, a "one-shot method" is usually the cheapest method for large quantities. In general, it is possible to obtain closer dimensional tolerances as we go from 1 to 4, but costs increase sharply (Fig. 1). The unit cost is, however, a function of volume and tooling costs, as well as of the method used. For very short runs where perhaps 25 developmental parts are needed, method 2 (machining unfired material) or 3 (machining partially fired material) may be cheaper than 1, since 1 may require a heavy investment in dies. Where more than 25 pieces are required and the shape of the piece is not too complex, it is probably cheaper to invest immediately in dies. Above 100 pieces, except in rare cases, the die method is cheaper. The numbers quoted refer to situations where the design is reasonably well frozen. Machining of blanks may be more desirable and less costly when it is anticipated that several design modifications will be required.

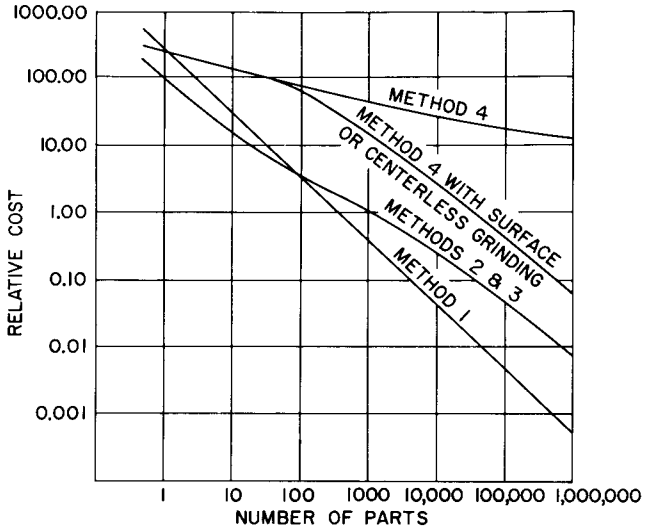


Figure 1. Influence of Forming Method and Volume of Pieces on Parts Cost

Method 2 when highly mechanized can be competitive with 1. In the spark-plug industry, for example, millions of high-quality ceramic parts are lathe-turned from unfired blanks at a cost of less than five cents per part. Methods 3 and 4 are ordinarily practical only when extremely critical dimensions (±0.0005 inch) or very low volumes (1 or 2 prototypes) are required. But there are important exceptions here also. Method 4, for example, can be very inexpensive for very high-volume production of simple shapes that can be handled automatically on a surface grinder or centerless grinder. (This statement does not apply to internal-grinding methods, however, which are very costly).

Forming Techniques Used to Make Finished Parts

The most common techniques for forming ceramics by the one-shot method (method 1) are pressing and extrusion. These will be explained more fully in the following pages. Injection molding and casting are also one-shot methods. The injection molding technique is capable of making extremely complex shapes but is not in wide use since the technique requires high quantities of plastic additives which makes it expensive and incapable of holding close tolerances. Casting is relatively slow and does not lend itself well to the high dimensional accuracy required by the electronics industry because the original shape contains large quantities of

water and shrinks excessively and nonuniformly during firing.

Forming Techniques Suitable for Subsequent Machining

All one-shot methods can be used to form blanks suitable for subsequent machining in the unfired, partially fired, or completely fired states as described in methods 2, 3, and 4. In addition, there are several specialized techniques which are only suitable when machining is employed. These are isostatic pressing and hot pressing.

In isostatic pressing, the ceramic powder is formed into a blank or "slug" of very uniform density. The material is formed by enclosing it in a rubber die or membrane which is subjected to hydraulic pressure. After pressing, the surface that was in contact with the rubber will be rough and irregular and will have to be shaped to the desired geometry by machining before or after firing. Because of the uniform density achieved by this method, control of dimensional tolerances and other properties is exceptionally good.

In hot pressing, the ceramic material is heated to a temperature near its softening point while it is being compacted in the die. This process is extremely expensive and complex at the present time. As a technique, it is used primarily to make very dense blanks from which a desired shape may be cut or machined; some pieces, such as rocket nozzles, are being pressed to the desired shape directly. This technique is being constantly improved, but it does not seem that it will be adaptable to high-volume, low-cost parts for some time to come.

In general, because of the limitations and costs of other methods, shapes and tolerances should be adapted to the one-shot method since the techniques of this method are, in most cases, capable of producing parts at high volume and low cost. To design effectively, the designer must be familiar with the processes of pressing and extrusion in detail. The pressing method is particularly well adapted to making relatively flat shapes, such as stems and spacers, and can be used to make shapes in which the height does not exceed the width or outside diameter. Extrusion is usually specified when long straight pieces are required.

Pressing Method

In the pressing method, the part is made in one operation by compacting a dry or slightly moist, granular powder in a carbide-lined or steel die. The simplest form of die, which is used to press a flat disk with no holes, consists of a round die cavity and flat-face top and bottom punches (Fig. 2).

In the first step of the pressing cycle, the die cavity is charged with the carefully prepared ceramic powder. The required weight of material for the part is obtained by controlling the position of the bottom punch. In the second step, the top punch enters the die and compacts the powder to a predetermined height. In the third step, the top punch retracts and the bottom punch moves upward, ejecting the formed piece. In the fourth step, the bottom punch drops to its original position and the cycle repeats.

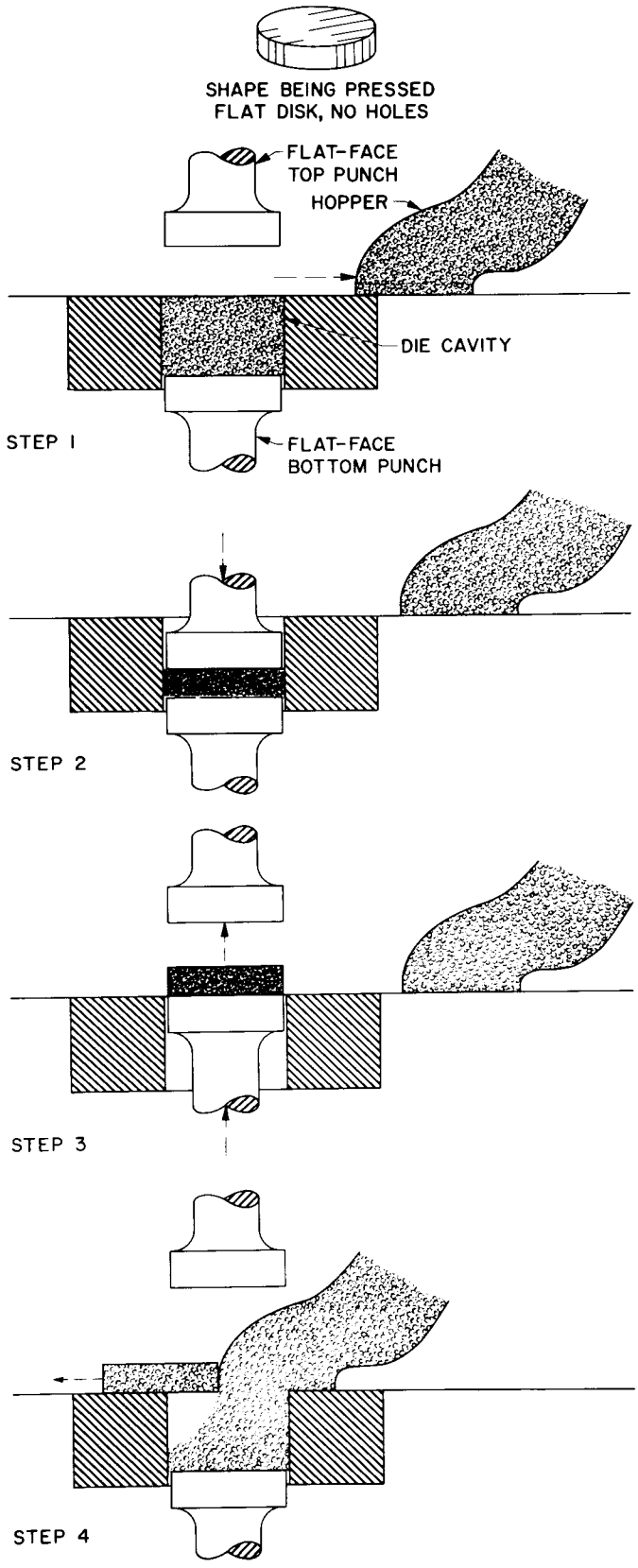


Figure 2. Steps in Pressing a Flat Disk

To make a similar disk with a hole (Fig. 3), it is necessary to add a core pin or pin bridge to the die. Alignment of the die is very critical since the pin passes through the lower punch and, in addition, engages a matching hole in the top punch during every compaction cycle. Wear is also a problem. Steel pins wear so rapidly that operation is limited to 100 to 25,000 parts depending on the composition of the ceramic being pressed. As a result, dimensional control is difficult. Carbide pins resist wear extremely well (500,000 to 1,000,000 parts are feasible) but are brittle and mechanically weak. Because small carbide pins are not practical, the designer should avoid specifying holes of less than 0.030 inch in pieces made of abrasive material like alumina. Smaller holes can be made, but only with steel pins. Because of the down-time required for frequent pin changes, however, the press output per hour will drop considerably and the costs of parts will be higher.

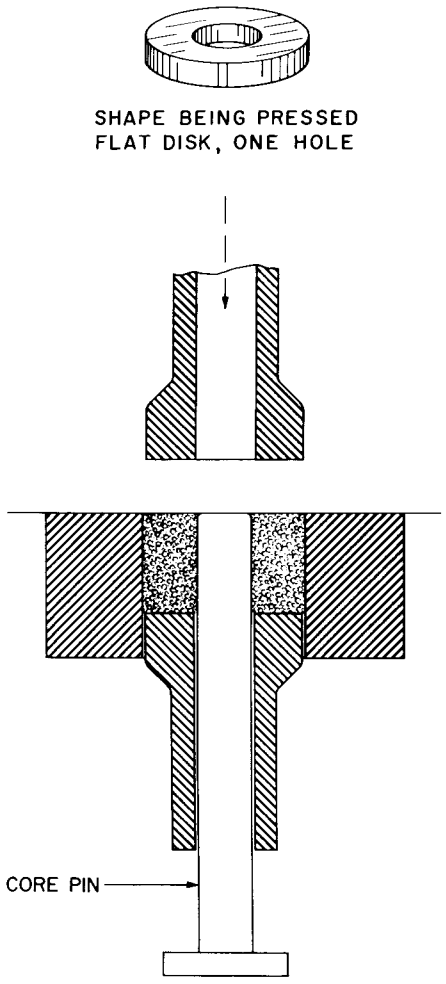


Figure 3. Die With Core Pin

To make a disk with shadow grooves (depressions that reduce electrical leakage), steps, embosses, or chamfered edges (Fig. 4), the faces of the punches must be contoured. Contouring adds appreciably to the initial tooling cost as well as to upkeep because the abrasive ceramic powders will cause the faces of punches to wear. To keep the tooling in good condition, contoured faces require expensive reshaping rather than simple surface grinding. Differences in thickness of the part also introduce forming problems. Laminations and cracks become more prevalent. In extreme cases, where the difference in step height is excessive, concentric sleeves, each of which can be controlled independently, must be used in place of one-piece punches (Fig. 5). Both the dies and presses needed to perform this type of work are more expensive, and the cost of the part increases.

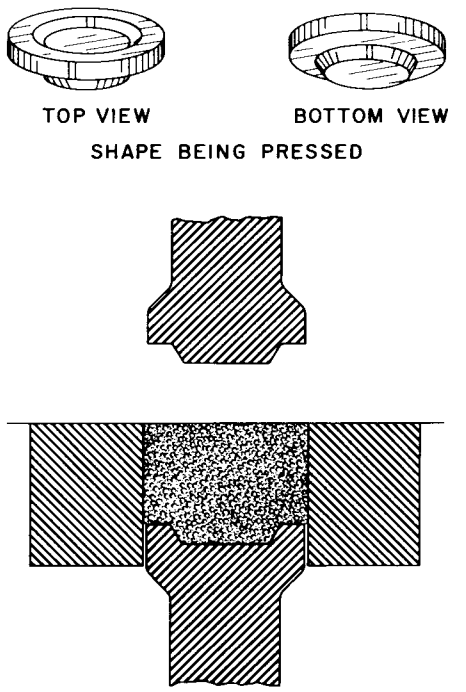


Figure 4. Die for Forming Disk With Shadow Grooves

The pressing method, in addition to being used for making disks or flat wafers, is sometimes used for thicker, chunkier shapes, such as blocks or cylinders. When the height of the part is nearly equal to or greater than the width or diameter, however, dimensional control is difficult. During the pressing of such parts, especially long, thin-walled cylinders, the powder in contact with the moving top punch is more highly compressed than is the powder in contact with the stationary bottom punch because of wall friction in the die cavity. This difference in density of the pressed piece influences the shrinkage which takes place during firing, and therefore, causes variations in dimensions.

To partially overcome the problem of nonuniform shrinkage, the ceramic supplier will often resort to a press in which both the top and bottom punch, and sometimes the cavity, move while the part is being compacted. By controlling the relative motions of the various sections of the die, it is possible to achieve greater uniformity and dimensional control.

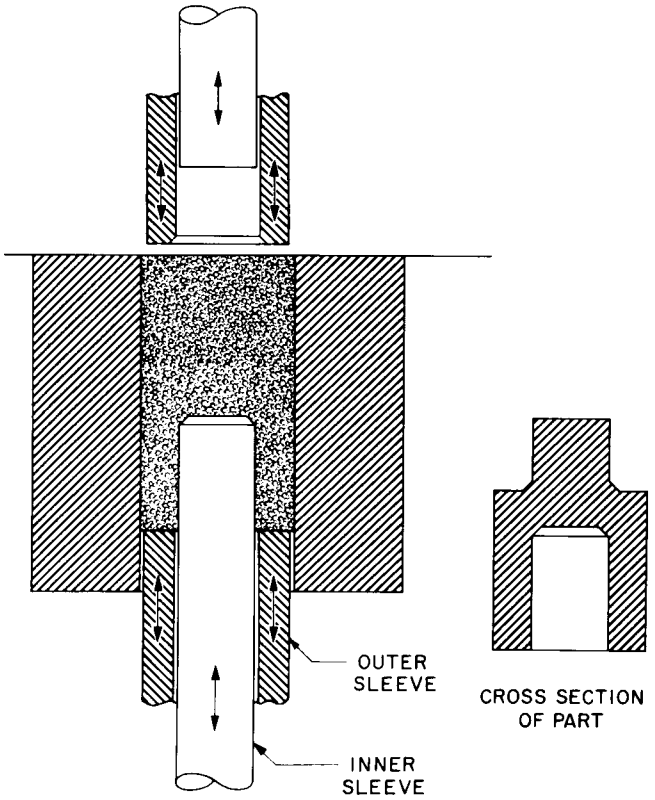


Figure 5. Die With Concentric Sleeves

Cylinders and other long parts made in a die where pressure is applied from only one end are characterized by a taper. The bottom, less-compacted portion will shrink the most and will be smaller than the top portion (Fig. 6).

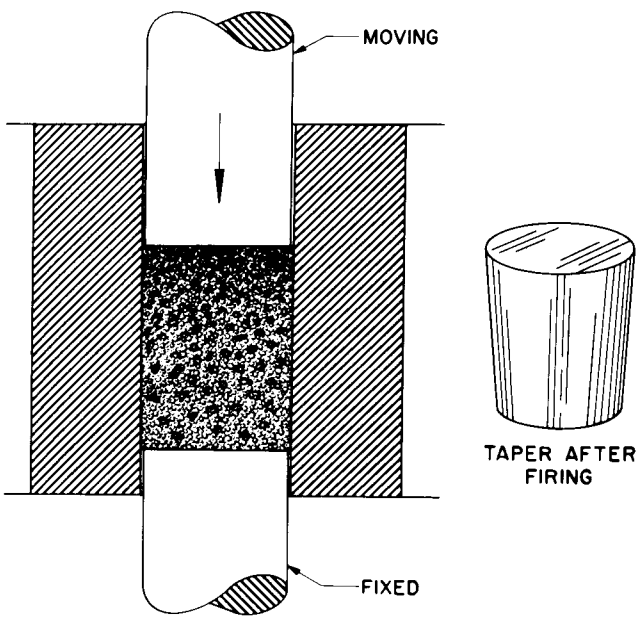


Figure 6. Shrinkage of Long Cylinder Pressed From One End

Parts pressed in properly designed multiple-motion dies should have acceptable tolerances. There are practical limitations, however, and parts pressed from both top and bottom often have slightly higher shrinkage in the middle than at the ends. This effect (Fig. 7) is called necking-in. The magnitude of the necking-in depends upon the "flowability" of the ceramic powder during compaction and varies with the composition of the ceramic being pressed. As a rough indication of the magnitude of the necking-in, a cylinder 1/2-inch high by 1/2-inch diameter with 0.050-inch walls will neck-in approximately 0.005 to 0.006 inch.

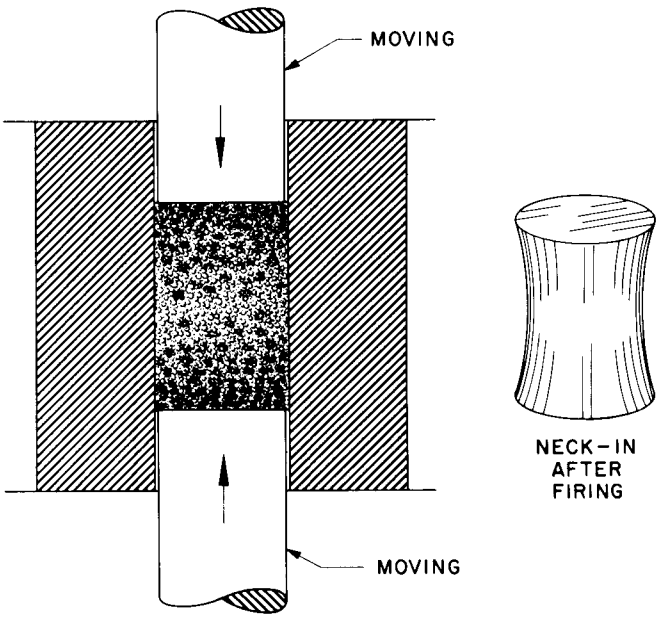


Figure 7. Shrinkage of Long Cylinder Pressed From Both Ends

From a description of the pressing process, it is obvious that holes cannot run through the piece at right angles to the height, unless they are put in by a separate operation. Holes parallel to the pressing axis cause little difficulty (except as already pointed out) although the more there are, the more the part costs. Dies which will make round holes or round parts are much more easily made than those for special shapes. For example, to make a square hole in the part, a square pin is required, and this in turn requires a square hole in the punch. The machining of these special pressing tools is a costly hand operation. Fig. 8 shows the complexity of tooling required for a relatively simple power-tube spacer.

There are other design recommendations which should be kept in mind in the design of pressed ceramic components. Under certain conditions some of these suggestions will have to be ignored. They should, however, be followed wherever possible since good de-

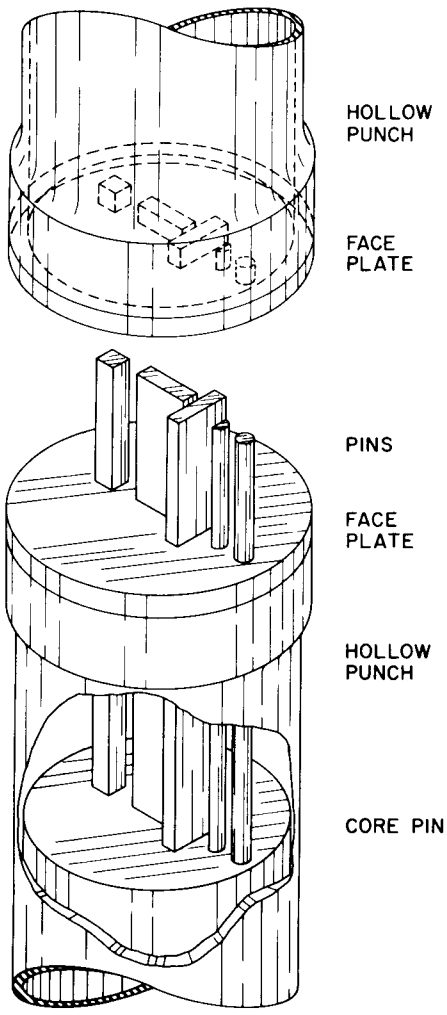


Figure 8. Typical Die For Pressing Tube Spacer

sign practice will result in lower parts cost. These recommendations are:

1. Allow maximum tolerances commensurate with good tube design. This requirement is particularly necessary for location of hole centers and diameters of holes. Where possible, use elongated holes for one of the grid side rods and one of the plate ears (Fig. 9A).
2. Design counterbores to permit about six degrees draft on the punch; draft permits faster pressing (Fig. 9B).
3. Do not place holes or counterbores so close together or so close to the edge that there will be excessive cracking during pressing, drying, or firing. The minimum safe distance between holes, or between holes and the edge of the part, is two times the hole diameter. A 0.030-inch hole, for example, should be at least 0.060 inch from the edge of the part (Fig. 9C). Parts are being made with distances less than this, but manufacturing difficulties are increased. If it is necessary that such holes or counterbores be in a position that may lead to large manufacturing losses, it is recommended that the wall be removed between the openings (Fig. 9D).
4. Avoid specifying holes normal to the extruding or pressing axis. Such holes must be made by drilling and a jig or template is required as a guide. If such holes must be specified, the part should be dimensioned from one end.
5. Avoid parts that require pressing dies of poor design, such as those with narrow splines or feather edges on the punches (Fig. 9E).
6. Avoid sharp internal angles (Fig. 9F) where possible. Such design leads to cracks in the part and to high die maintenance costs, both of which are finally reflected in higher cost of the part.
7. Avoid reentrant grooves or undercuts (Fig. 9G) where possible, since such grooves must be made by machining.
8. Make parts as nearly symmetrical as possible; avoid large, abrupt differences in thickness (Fig. 9H) or unequal cross sections (Fig. 9I).
9. Visualize the metal tool which will be used to make the ceramic part. Of all the design criteria, this one is perhaps the most important because it can help to clarify the designer's thinking. To make a ceramic part with a number of holes near the edge, the toolmaker must first be able to make a metal punch with holes near the edge. To make a ceramic cylinder with a fluted inner surface, the toolmaker must first be able to make a pin with the opposite contour. By reflecting on these and similar problems, the designer will arrive at more reasonable and practical ceramic shapes. In essence, don't expect to obtain ceramic shapes and tolerances which cannot be translated into the metal shapes and tolerances necessary to make a workable die.

Extrusion Method

In the extrusion method, a rather soft, pasty mass is forced through a die nozzle, much as toothpaste is squirted from a tube. These long lengths are dried, often (to minimize warpage) in a special humidity-controlled dryer, after which the pieces are cut to length and then fired. Some materials are fired first and then cut to length.

Many of the same design principles for pressed parts apply to extruded ware. Symmetry of cross section (Fig. 9I) is even more important, however, to the achievement of freedom from warpage and nonuniform shrinkage during drying and firing. Unless additional handling and machining can be tolerated, the ends of the pieces must be flat because pieces are cut to length.

The ratio of the wall thickness to the diameter of extruded pieces is very important. Concentricity of the inside and outside diameters and out-of-roundess are always a problem with extruded work, but can be controlled if care is taken. As a rule of thumb, wall thickness should be at least equal to one-tenth of the outside diameter.

TOLERANCES

Standard industry tolerances are ± 0.005 inch or ± 1 per cent, whichever is greater. This rule means, for

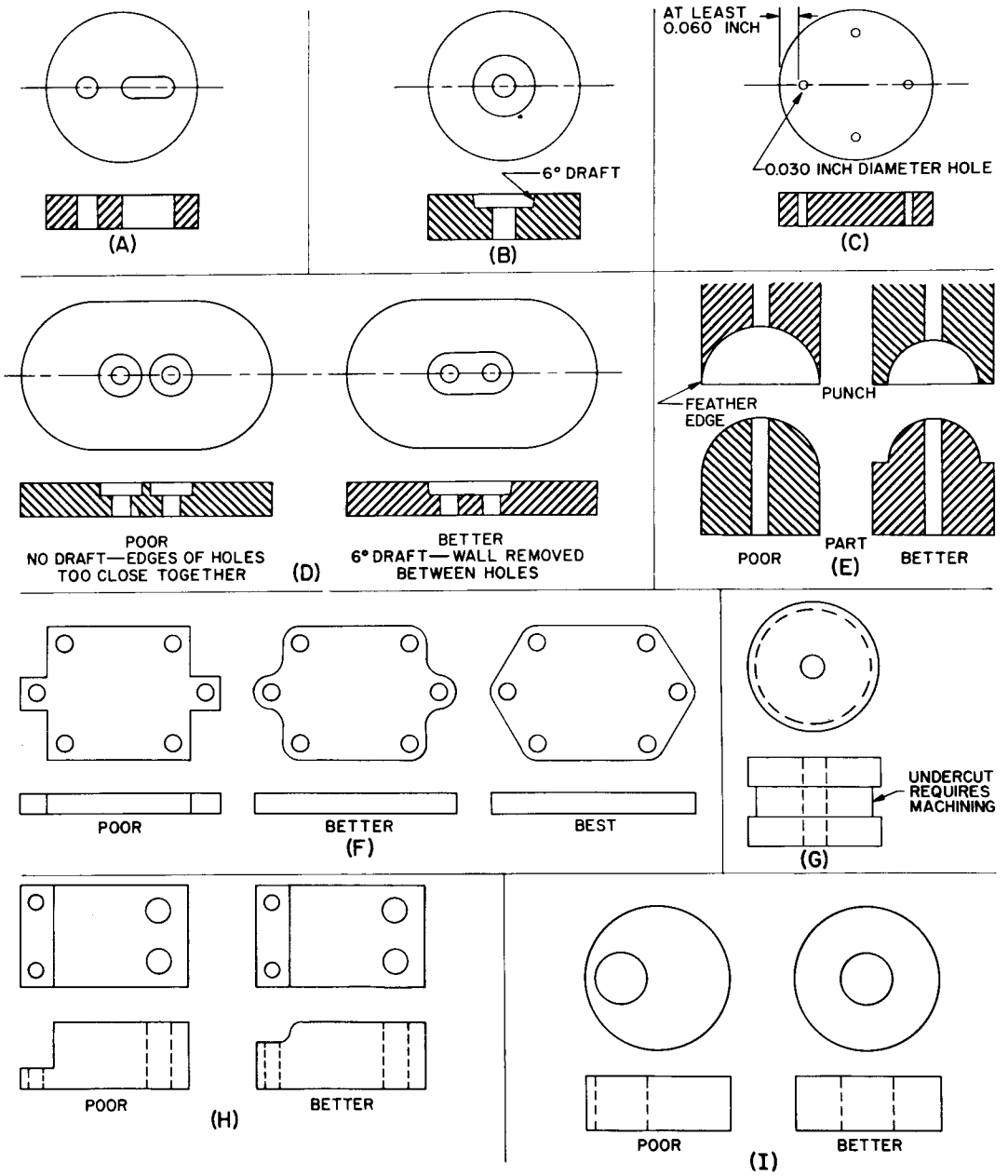


Figure 9. Examples of Good and Bad Design Practices

example, that dimensions from zero up to 0.500 inch will not be held to better than ±0.005 inch. Above 0.005 inch, the ±1 per cent tolerance is in effect; i.e., ±0.010 inch on a 1.000-inch part. Better tolerances can be obtained, of course, but at an additional cost.

The ceramics manufacturer hesitates to guarantee better dimensional control because ceramic pieces shrink in size during firing. Many ceramic compositions shrink as much as 20 per cent. * To obtain a part 1.000 inch in diameter, the ceramist must press a 1.200-inch piece. The linear shrinkage which occurs during firing is affected by small variations in compo-

*Per cent linear firing shrinkage may be calculated on either the pressed or fired basis. The most convenient method is to use the fired basis:

$$\frac{\text{Pressed Diameter - Fired Diameter}}{\text{Fired Diameter}}$$

x 100. This method permits the ceramist to multiply a finished part design by a factor of 1 + the per cent shrinkage to obtain the unfired part dimension (die size). In the example cited, the designer specifies a 1.000-inch part. The ceramist multiplies this by 1.200 to obtain 1.200 inch as the required dimension of the pressed piece before it is fired.

In reverse: $\frac{1.200\ (\text{Pressed}) - 1.000\ (\text{Fired})}{1.000\ (\text{Fired})} \times 100 = 20$ per cent shrinkage.

sition, powder-mixing technique, degree of compaction, die wear, and firing schedule. As a result, exceptional control is required to obtain tolerances of a few thousandths of an inch on the finished part.

In specifying dimensional tolerances the designer should be familiar with the principles involved. Knowing these, he can often obtain better than standard tolerances even from commercial suppliers.

1. The lower the firing shrinkage of a particular composition, the better the tolerances which can be held.
2. The smaller the part (or dimension involved), the easier it is to hold tolerances.
3. The tighter the limits on big dimensions of the part, the closer will be the tolerances on small dimensions.

These principles must be applied intelligently and with discretion. Because dimensional control is largely a function of the amount of firing shrinkage, it is obvious that the designer should specify a ceramic composition having the lowest shrinkage factor consistent with other property requirements. Shrinkage figures are not published, but most ceramic vendors will be glad to work with the designer if he will tell them his needs.

Suggestion 2 points out that shrinkage is a percentage of the over-all dimension. If the firing shrinkage for a particular material and process is exactly 20 per cent, a 1.200 inch ± 0.000 part will fire to exactly the required 1.000 inch ± 0.000. A 0.024 inch ± 0.000 part will shrink to exactly 0.020 inch ± 0.000. But suppose, as is more usual, the shrinkage varies slightly. If it were 21 per cent instead of 20 per cent the 1.200-inch part would fire down to 0.992 inch which is 0.008 inch undersize. The 0.024 inch part, however, would fire to 0.0198 inch, only 0.0002 inch undersize.

Look at this another way. How much can the shrinkage vary from 20 per cent if it is necessary to hold ± 0.005-inch tolerances on both a 1.000-inch part and a 0.020-inch part: For the one-inch part, it would be necessary to hold shrinkage to between 19.5 and 20.5 per cent, whereas shrinkage for the 0.020-inch ± 0.005 part could vary between zero and 60 per cent.

Suggestion 3 is a corollary of 2. It has already been shown that holding a 1.000-inch dimension to a fairly reasonable ± 0.005-inch tolerance automatically holds a 0.020-inch dimension in the same part to less than ± 0.0002-inch tolerance. Care is required, however, in the use of this technique. If the 0.020-inch dimension is a hole size, wear on the pin which forms the hole in the pressed piece could cause a decrease in size independent of processing and material variables. To control this it would be necessary to use long-wearing pins and to check their dimensions regularly.

It is important that the designer also be aware of the reverse implications of suggestion 3. Except to control die or pin wear, it is pointless to put very close tolerances on only one dimension of a part, thinking that the part will be cheap because the other dimensions are not critical. Shrinkage takes place uniformly throughout a properly designed and formed piece, and basically, the cost is determined by the most critical dimension.

There is one important exception to this rule. Height tolerances are, in a sense, independent of width or diameter tolerances. It is reasonable, therefore, to specify close tolerances on diameter and fairly loose tolerances on height, with attendant savings in cost. This distinction is, in fact, desirable because most dimensions are a function of the die design and are "built-in," whereas the height is controlled only by the precision and setting of the press itself, i.e., the variability of the distance to which the top punch descends into the die cavity during each stroke.

The art of designing practical, high-precision, low-cost ceramic shapes is not as difficult nor as complex as it may seem. It does, of course, require some experience and judgment. Intelligent cooperation and mutual understanding between the designer and ceramic supplier can often reduce the problems of both.

CHOICE OF MATERIAL

Usually the choice of a particular ceramic composition is based not only on the cost or ease of forming the ceramic, but on some critical property of the material such as its strength, thermal conductivity, or thermal expansion. Fortunately, there are many ceramic compositions available with a wide variety of properties. In this chapter and in the listed references typical properties are given for a number of these ceramic materials.

As has already been pointed out, the cost of a ceramic part is a function of the forming technique, tolerance requirements, and number of pieces required. In addition, cost is a function of the ceramic composition used to obtain the desired properties. It is cheaper, for example, to make parts from steatite than from alumina. There are several reasons for this: (1) the individual raw materials from which steatites are made are widely available at low cost in a form suitable for ceramic use, whereas suitable aluminas are most costly and often require additional processing, such as ball milling, by the user; (2) the mixing equipment and presses used to process and fabricate alumina compositions are more expensive because alumina is harder and more abrasive than steatite; (3) kilns and kiln furniture for firing alumina pieces are more expensive than for steatite because alumina is more refractory and must be fired at higher temperatures to achieve a dense, gas-tight part.

Similar factors determine, to some degree, the cost of parts made from other ceramic compositions. It is difficult, however, to make completely valid, broad generalizations as to cost because new sources of raw material, improved forming techniques, and increased demand for a given composition often are more important in establishing cost than are the factors just discussed.

In evaluating ceramics having properties suitable to a given application, the designer will encounter names

like "steatite," "forsterite," and "alumina." It is extremely important to recognize that these names are generic. That is, they refer to families of materials. They do not apply to single, specific compositions.

Steatite, for example, refers to a group of compositions which are reasonably similar in properties. These properties vary from vendor to vendor, however, since each manufacturer develops his own blend of raw materials and uses his own processing techniques. Unless the vendor is specified, therefore, it is impossible to give exact values of properties.

It is for this reason that substitution of an apparently similar material made by one manufacturer for that made by another is not always successful. Before specifying an alternate vendor, it is necessary to compare carefully the properties listed by both suppliers and to make comparison tests on actual assemblies.

The same precautions are equally desirable when a new part is to be designed. There are ten or twenty sources, for example, of high-strength vitreous alumina bodies. In addition, each of these sources may be able to supply three or four modifications. Although, as a class, all of these alumina ceramics will have property values which distinguish them from steatite, one alumina ceramic may be more resistant to thermal shock than a second. A third, which is intermediate in this respect, may surpass both of the others in electrical characteristics.

Since there is, as yet, little standardization of ceramic compositions and properties from vendor to vendor, it is the designer's responsibility to choose wisely. He must determine which properties are most important in a given application, and he must then select the material and vendor best able to meet and maintain the necessary requirements. Vendor's catalogs and other published material go out of date very quickly. The designer should maintain an active file of current literature and should use this chapter as background material to aid in the correct interpretation of the manufacturers' data. Only by such care can he be assured that he will obtain the best possible choice of properties at the best possible price.

Because many of the terms used in the property charts are unfamiliar to the tube designer, the following sections have been included to help the designer to evaluate competing materials or products more easily.

MECHANICAL PROPERTIES

Strength

There are three common ways of expressing strength: compressive strength, tensile strength, and modulus of rupture (or flexural strength).

All ceramic materials are relatively weak in tension. Every effort should be made to use them in a way such that they are exposed only to compressive forces. Although materials vary, the compressive strength is usually 10 to 20 times greater than the tensile strength. In practice, however, these high values are never reached because compressive loading forces are altered by shear and twist forces and the ceramic, therefore, fails in tension.

It is difficult to get good compressive and tensile data because the test shapes and the testing equipment are quite complex. Modulus of rupture figures are more commonly quoted. To obtain these values, a bar or rod is supported at the ends, and the load is applied in the middle. The value for modulus of rupture, which presumably takes into account the span, the load at rupture, and the cross-sectional area, is calculated. This modulus is supposed to be representative of the strength of the outer "fibers" of the ceramic, which are exposed to tension. Modulus of rupture values are actually about 50 to 70 per cent of the actual tensile strength.

The designer should realize that strength values can not be taken literally. Even though the formula used to calculate them takes into account variations in sample size, in actual practice the values obtained are influenced significantly by the size of the specimen.

Results can also be misleading because of variations in the surface condition of the sample. By "quenching" the sample so that the skin is in compression, higher values can be obtained. Samples cut from a larger piece, or ground to size, will normally give lower values because the "fired" skin has been removed. Careful grinding and polishing which remove surface flaws, however, increase the strength.

Strength tests are essentially a measure of flaws or weak spots and, therefore, reproducibility of values is a function of the skill with which the ceramic pieces are prepared. Strength figures, to be meaningful, should be based on a minimum sample of 20 pieces.

Some persons consider that the lowest strength value determined during a test is the important figure since there is always a chance that the device may contain some pieces which will fail at this value. Others prefer to choose a material which has a narrow range with no low values even though the average may be less than that for a material having a wider range of values. Still others select materials on the basis of the highest single value and then direct their efforts toward improving the process or inspection procedures to eliminate the weak, flaw-ridden pieces.

Where strength is the dominant reason for selecting a particular ceramic, the simplest and best method is to choose the ceramic having the highest nominal strength consistent with other required properties. Samples should be obtained and tested by an arbitrary but reproducible load test. After the proper values are correlated with good performance in the device, all future incoming lots or substitute materials should be tested against this standard.

When an acceptable material and shape suddenly go "sour" for strength, the most common cause is a change in the forming or firing technique used by the manufacturer. These changes, in turn, may be due to va iations in the chemical and physical properties of the raw

materials. The easiest way to determine if such changes are the cause is to compare microstructures with previously established reference samples of good and bad parts. Grain size, pore structure, and amount of glass should be noted. Changes in the microstructure often correlate with changes in fired density; as a result, density measurements are a useful tool for control purposes.

At the present time (1960), vitreous alumina ceramics are outstanding when strength is the main requirement. These materials have been improved substantially during the past ten years. Original flexural strength values of approximately 30,000 pounds per square inch have been pushed upward; today, values of 50,000 to 60,000 pounds per square inch are commonplace.

Table I gives the flexural strengths of a number of materials. It is important to remember that the mechanical and physical properties of a given type of ceramic will vary within wide limits between suppliers, or even between grades of a given supplier, depending upon the raw materials used and the treatment they have received. For that reason it is impractical to quote a single numerical value for any property.

TABLE I

Flexural Strengths of Ceramic Materials
(listed in order of decreasing strength)

Type	Strength-lbs/in^2
Alumina	8000-62,000
Zircon	9300-26,500
Forsterite	19,000-25,000
Steatite	11,000-22,000
Porcelain	6760-11,500
Lava	9000
Cordierite	6140-8000

Impact Strength

Impact strength is normally of little importance to the tube designer. It was originally used to describe the resistance of porcelain dinnerware to handling and chipping. The test consists of supporting a rod at one end, cantilever style, and striking it with a metal bob which swings through an arc, pendulum fashion. The bob is released from higher and higher points on the arc until the rod breaks. The higher the figure reported, the higher the impact strength. The units are usually reported in inch-pounds, but sometimes arbitrary units of height or length of arc are used.

Specific Gravity, Porosity, and Water Absorption

Specific gravity, porosity, and water absorption are very closely interrelated. Every ceramic composition has an ultimate, theoretical fired density (specific gravity) which is approached commercially, but seldom attained. The difference between the apparent fired density and the theoretical fired density is caused by pores and voids in the fired piece.

The total porosity of a piece includes both the open pores, which connect to the surface, and the closed pores, which do not. It is possible for a material which does not fire to theoretical density to be classed as a dense, vitreous, gas-tight ceramic if all the pores are closed pores.

Water absorption is really a measure of the open pores.

The terms are interrelated in the following fashion:

Per cent open porosity = per cent water absorption x apparent fired density because:

$$\frac{V_{H_2O}}{V_F} = \frac{W_{H_2O}}{W_F} \times \frac{W_F}{V_F}$$

since

$$V_{H_2O} = W_{H_2O}$$

where

W_F = Fired weight of ceramic

V_F = Fired volume of ceramic

V_{H_2O} = Volume of absorbed water

W_{H_2O} = Weight of absorbed water

Also: per cent total porosity = 100 -

$$\left(100 \times \frac{\text{apparent fired density}}{\text{theoretical fired density}}\right)$$

For use inside a vacuum tube, a part should either be vitreous (no open pores) or very porous (about 30 per cent). Degassing of vitreous parts is easy because only gas adsorbed on the external surface must be removed. Degassing parts of high porosity is easy because there are many large interconnected open pores which can be pumped out rapidly. Parts with low porosities (1 to 10 per cent) are dangerous because the pores are partially blocked, or are of small cross section, and therefore cannot be pumped out completely in a short time. Such parts give off gas gradually on life and poison the cathode.

For parts which must be vacuum tight, water absorption values based on weight measurements are not sensitive enough. The most reliable method is to put the parts in a pressure bomb filled with a solution of methyl violet dye and to subject them to a pressure of 20,000 pounds per square inch overnight. After the parts are removed, they are broken open and examined. A gas-tight part will show no evidence of dye penetration.

THERMAL PROPERTIES

Maximum Operating Temperature

In specifying ceramic parts, especially those used in seal and brazing applications, it is important to know the temperature to which a part can be exposed before it deforms and loses its dimensional accuracy. Although information on safe operating temperature or

on softening temperature is often given in tables, these values can be used only as a guide since the shape and mass of the piece, the amount of load it must support, and the length of time it is to be held at high operating temperatures affect the results. Ceramic parts should never be used at temperatures higher than 50 to 100 C below the firing temperature used by the supplier.

Properties of ceramics are a function of the composition of the ceramic and the processing temperatures. Extended use of ceramic parts at temperatures near the original firing temperature may result in continued reaction between the constituents with a resultant change in the properties of the product.

In general, since ceramics are mixtures of crystal and glass phases, they behave like glass and soften gradually rather than melt as the temperature increases. With pure oxides or eutectic compounds, however, the melting point may be quite sharp. Often a good indication of the deformation temperature can be obtained from the phase-equilibrium diagrams published by various technical societies.

Sometimes two compositions, each quite refractory, react when in contact to form a material having a melting point lower than that of either of the compositions. Again, the phase diagrams are helpful in predicting or interpreting this phenomenon.

The atmosphere to which the heated part is exposed also exerts an influence. Most tables of melting points and maximum safe operating temperatures are based on use of the materials in oxidizing atmospheres. Values in a reducing atmosphere or vacuum are generally lower, but accurate data are scarce.

Selection of a material based on safe operating temperature is very seldom a problem. The ceramic compositions likely to be selected by the designer for some special property, such as strength or thermal expansion, are generally capable of withstanding temperatures far in excess of those encountered in the manufacture or use of receiving tubes. Melting points and ranges of safe operating temperatures are given in Fig. 10.

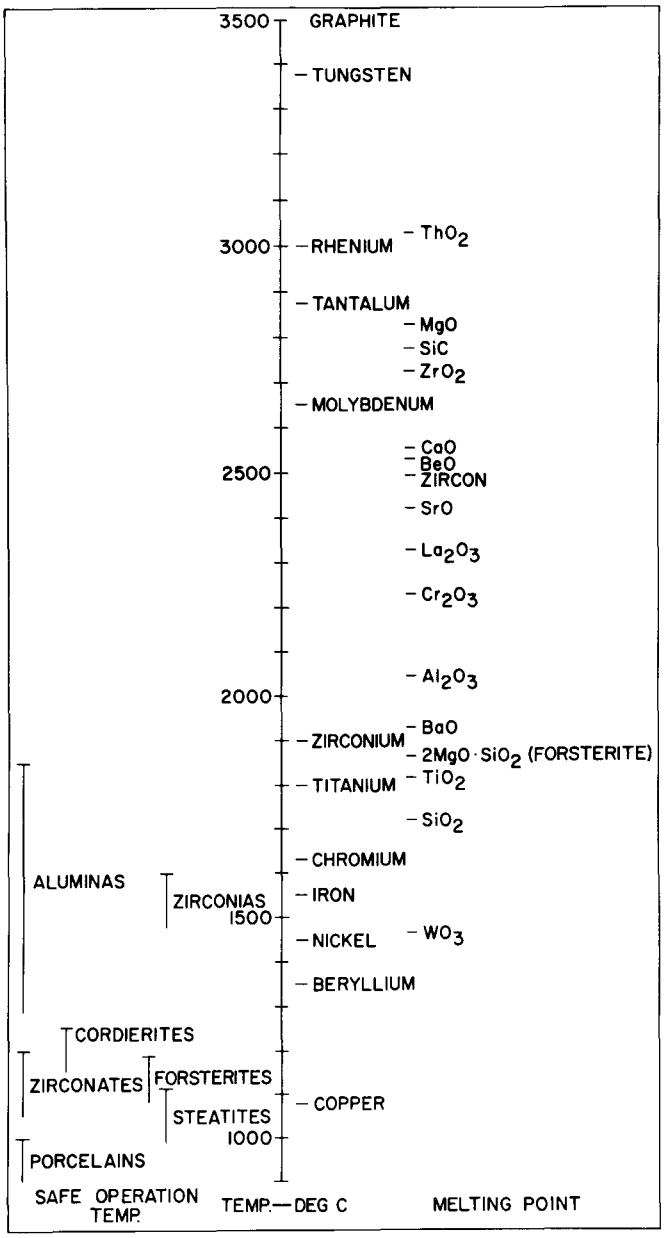


Figure 10. Melting Points and Safe Operating Temperatures of Ceramics and Metals

Thermal Conductivity

Thermal conductivity is primarily a function of the composition and porosity of the fired ceramic. It also varies with temperature. The temperature coefficient can be positive or negative. In crystals the conductivity is high at low temperature and decreases with rising temperature. In glasses and "amorphous" solids, the reverse is generally true. In polycrystalline ceramics, the highest conductivity is achieved by having low porosity, large grains, and a minimum of glass.

Compared with metals, ceramics have low thermal conductivities. An exception is beryllia which, at room temperatures, has a thermal conductivity equal to that of aluminum. Because of beryllia's toxicity and high melting point, it is costly and difficult to fabricate and, therefore, is limited in use.

The thermal conductivities of silicon carbide and magnesia are high, but they are considerably below that of beryllia. Silicon carbide is an electrical conductor and has limited use in tubes. Magnesia (periclase) is a good electrical insulator but, because of its high melting point, is difficult to fabricate into a dense, vacuum-tight part. It is also relatively unstable and reacts slowly with moisture from the air. It has limited use as a spacer to be used in vacuum tubes because it decomposes in contact with hot tungsten. Of the more practical, available materials, alumina has the highest thermal conductivity.

Thermal conductivity data are reported in a variety of units. Table II gives useful conversion factors. Multiply the "units given" by the factor given to get "desired units."

TABLE II

Conversion of Units of Thermal Conductivity*

Units Given	Conversion Factor for Desired Units			
	$cal \cdot sec^{-1} \cdot cm^{-2} \cdot cm \cdot deg\ C^{-1}$	$watt \cdot cm^{-2} \cdot cm \cdot deg\ C^{-1}$	$Btu \cdot hr^{-1} \cdot ft^{-2} \cdot in \cdot deg\ F^{-1}$	$Btu \cdot hr^{-1} \cdot ft^{-2} \cdot ft \cdot deg\ F^{-1}$
$cal \cdot sec^{-1} \cdot cm^{-2} \cdot cm \cdot deg\ C^{-1}$	1	4.1814	2903	241.9
$watt \cdot cm^{-2} \cdot cm \cdot deg\ C^{-1}$	0.2390	1	693.8	57.78
$Btu \cdot hr^{-1} \cdot ft^{-2} \cdot in \cdot deg\ F^{-1}$	3.445×10^{-4}	0.001441	1	0.08333
$Btu \cdot hr^{-1} \cdot ft^{-2} \cdot ft \cdot deg\ F^{-1}$	0.004134	0.01730	12.0	1

* After Ref. 11

TABLE III

Thermal Conductivities *

of Pure, Zero-Porosity, Polycrystalline Oxide Materials

Material	$Watts \cdot Cm^{-2} \cdot Cm \cdot Deg\ C^{-1}$ at Specified Temperature (deg C)									
	100	200	400	600	800	1000	1200	1400	1600	1800
Al_2O_3	0.302	0.225	0.132	0.0913	0.0720	0.0616	0.0553	0.0549	0.0607	(0.0757)
BeO	2.195	1.745	0.930	0.469	0.270	0.203	0.172	0.164	0.151	(0.154)
CaO	0.152	0.111	0.0917	0.0830	0.0804	0.0780	---	---	---	---
Forsterite	0.0539	0.0452	0.0358	0.0290	0.0268	0.0244	0.0237	(0.0233)	---	---
Graphite	1.782	1.495	1.121	0.921	0.762	(0.625)	(0.549)	---	---	---
MgO	0.359	0.282	0.165	0.115	0.0850	0.0700	0.0612	0.0603	0.0688	(0.0946)
Mullite	0.0612	0.0552	0.0473	0.0431	0.0406	0.0397	0.0387	(0.0387)	---	---
NiO	0.124	0.0994	0.0716	0.0570	0.0460	0.0449	---	---	---	---
Spinel	0.149	0.129	0.102	0.0812	0.0665	0.0577	0.0544	---	---	---
TiO_2	(0.0654)	0.0500	0.0391	0.0361	0.0339	0.0330	0.0330	---	---	---
ThO_2	0.1025	0.0855	0.0599	0.0435	0.0340	0.0305	0.0251	(0.0245)	---	---
Zircon	---	0.0570	0.0570	0.0465	0.0432	0.0409	0.0393	0.0383	---	---
ZnO	---	(0.171)	0.112	0.0700	0.0549	---	---	---	---	---
ZrO_2	0.0195	0.0196	0.0205	0.0209	0.0219	0.0228	0.0239	0.0244	---	---

* After Ref. 10
Values in parentheses are extrapolated

Many thermal conductivity data are available in the literature; they are continually being refined and modified. Fig. 11 gives only typical values believed to be of interest to the designer. Commercially available ceramic bodies differ in composition, grain size, glass content, and porosity from those covered by Table III, and it is necessary to consult the manufacturers' property charts for specific data.

For a given material, the change in thermal conductivity due to porosity is essentially linear at temperatures up to 500 C. A ceramic having a porosity of 10

per cent, for example, will have a thermal conductivity which is 90 per cent of the absolute value given for zero porosity. Above 500 C, the size, shape, and orientation of the pores, and the emissivity of the material become important because radiation, as well as conduction, takes place. Loeb[10] has developed an expression which takes these effects into account.

$$K_p = K_s \left[(1 - P_c) + \frac{P_c}{\frac{P_L K_s}{4\sigma\epsilon\gamma\ dT_m^3} + (1 - P_L)} \right]$$

where K_p = conductivity of porous sample
K_s = conductivity of solid sample
P_c = cross-sectional pore fraction
P_L = longitudinal pore fraction
δ = Stefan radiation constant
γ = geometrical pore factor
= 1 for laminar pores and cylindrical pores with axes parallel to heat flow direction
= 2/3 for spherical pores
= $\pi/4$ for cylindrical pores with axes perpendicular to heat-flow direction
ϵ = emissivity of pores
d = dimension of pore in direction of heat flow
T_m = mean absolute temperature

Fig. 11 combines pertinent data from Refs. 8 to 11 recalculated into watts·cm^{-2}.cm·deg C^{-1} for easier use and comparison. The designer will note that the bulk of the ceramic compositions fall toward the low end of the scale.

Specific Heat

The specific heat of a material is the quantity of heat required to impart a unit increase in temperature to a unit mass of the substance and is commonly expressed in calories per degree centigrade per gram (cal/deg C/gm). Some writers introduce a dimensionless specific heat by defining it as the ratio of the quantity just defined for the material to the corresponding quantity for water. Values for various ceramics are shown in Fig. 12.

Thermal Expansion

The thermal expansivities of ceramics are generally lower then those of metals. The expansion of most well-fired stable oxides or ceramic compositions is reasonably linear, although the curves normally steepen gradually as the temperature is increased. Oxides such as silica and zirconia, or compositions containing these oxides in unreacted form, have erratic curves (Figs. 13 and 14). These rapid changes in expansion at certain temperatures are caused by inversions from one crystalline form to another.

The thermal expansion of silica can be stabilized by fusing it into a glass, but zirconia is too refractory and does not readily form a glass. Instead, zirconia is stabilized by adding 2 per cent lime or magnesia (Fig. 15). Both silica and zirconia are common additions to ceramic compositions. When they are combined with other substances to form new compounds such as mullite ($3Al_2O_3 \cdot 2SiO_2$) or zircon ($ZrO_2 \cdot SiO_2$), they lose their identities.

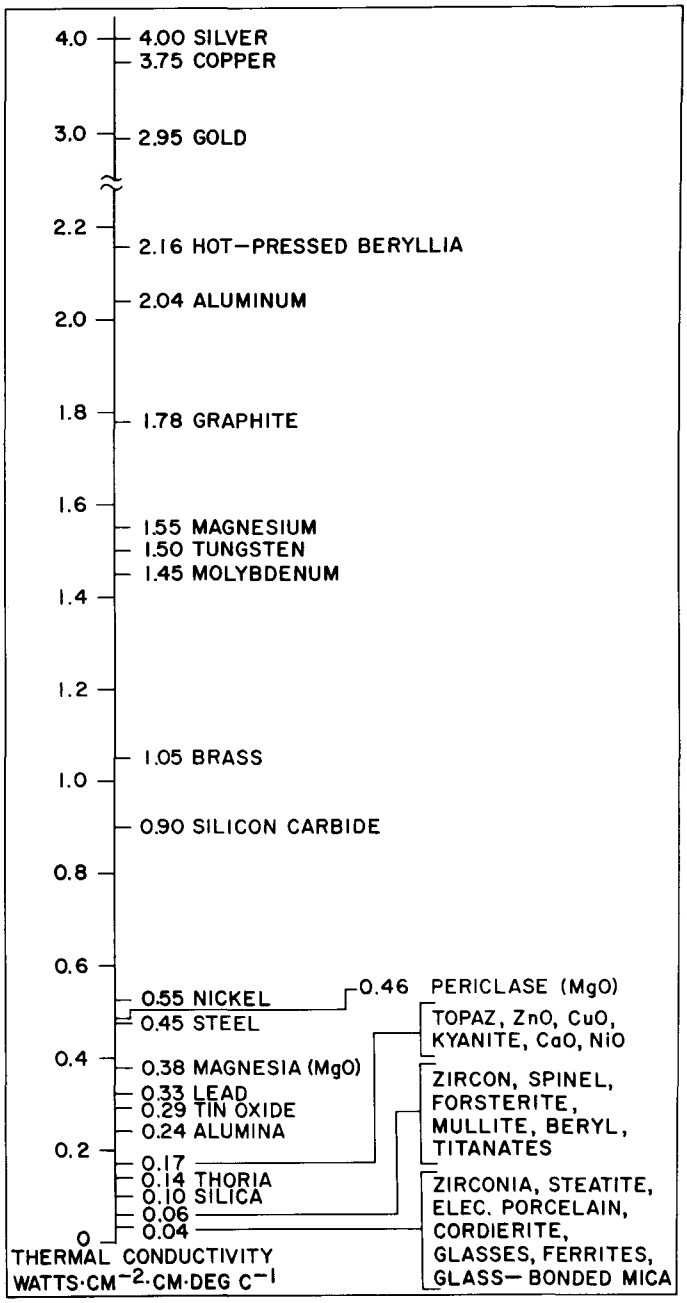


Figure 11. Thermal Conductivities of Ceramics and Metals

Among the available ceramic compositions, the designer will find materials having thermal expansions ranging from 13.0 x 10^{-6} down to some which are zero or negative.

Expansion coefficients are usually given in inches per inch per degree centigrade (in./in./deg C) for a specified temperature or temperature range. Sometimes they are given as per cent expansion. To convert from

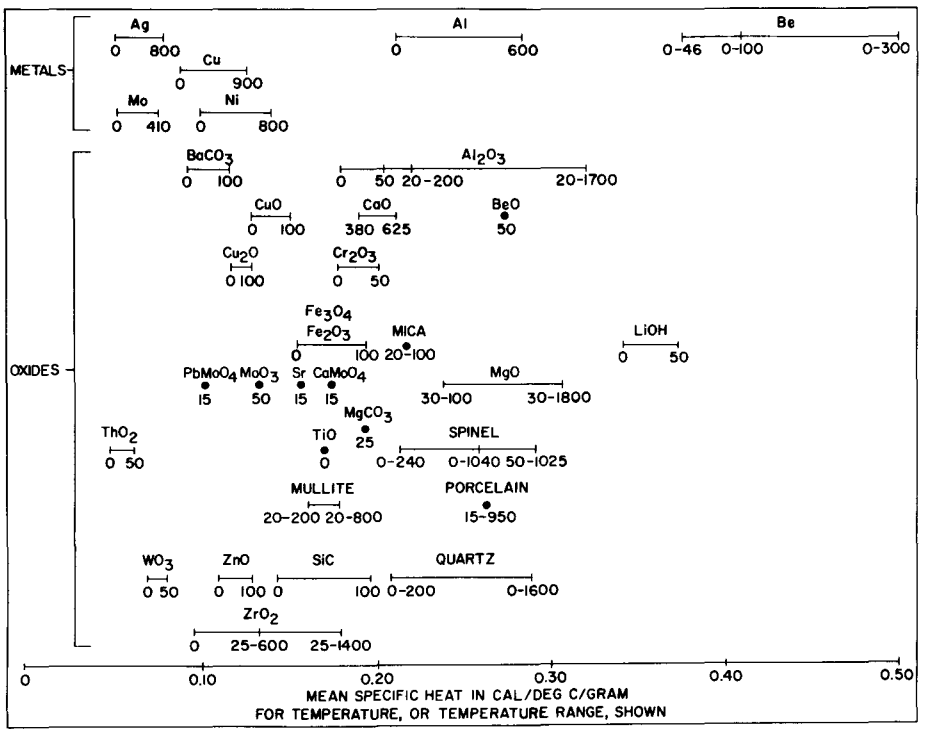


Figure 12. Specific Heats of Ceramics and Metals

per cent expansion at a given temperature to the coefficient, the per cent expansion is converted to inches and is then divided by the temperature.

Example: a graph shows magnesia (MgO) to have an expansion of 0.56 per cent at 500 C. This value is equivalent to 0.0056 inches per inch. Divide by 500 C. Then,

$$\frac{0.0056 \text{ in.}}{(1.0000 \text{ in.})(500 \text{ C})} = \frac{5.6 \times 10^{-3} \text{ in.}}{5 \times 10^{2} \text{ in. C}} =$$

$$1.12 \times 10^{-5} \text{ in./in./deg C (at 500 C)}$$

This coefficient is more commonly written 11.2×10^{-6}/deg C.

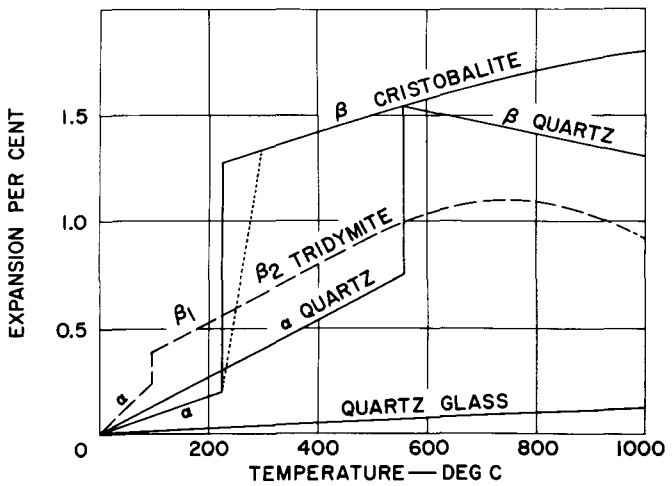


Figure 13. Expansion of Various Forms of Silica as a Function of Temperature

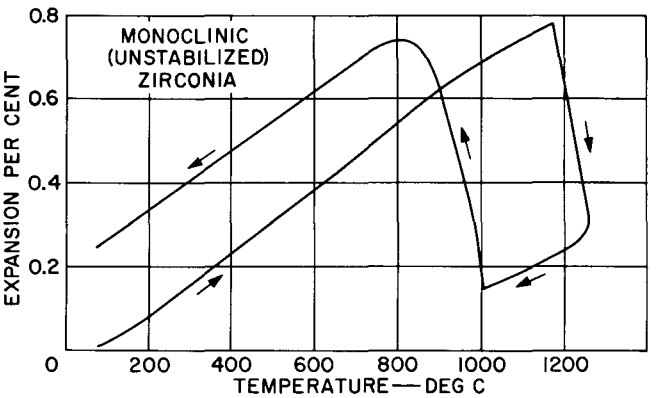


Figure 14. Expansion of Monoclinic (Unstabilized) Zirconia as a Function of Temperature

To convert the coefficient into per cent expansion, multiply the number of degrees by the coefficient to obtain the increase in length per unit length. Multiply by 100 to convert to per cent.

Expansion values given for single-crystal materials, such as single-crystal sapphire, may be different along each crystallographic axis. In single-phase polycrystalline compositions such as sintered alumina, the individual crystals are randomly oriented and the expansion will be an average of the values along each single-crystal axis. Polycrystalline ceramic compositions that consist of two or more phases which differ in thermal expansion will have an expansion which is roughly a function of the amount of each phase present according to the following relationship:

$$\text{Coef (A + B)} = \frac{(\text{Coef A} \times \%\text{A}) + (\text{Coef B} \times \%\text{B})}{100}$$

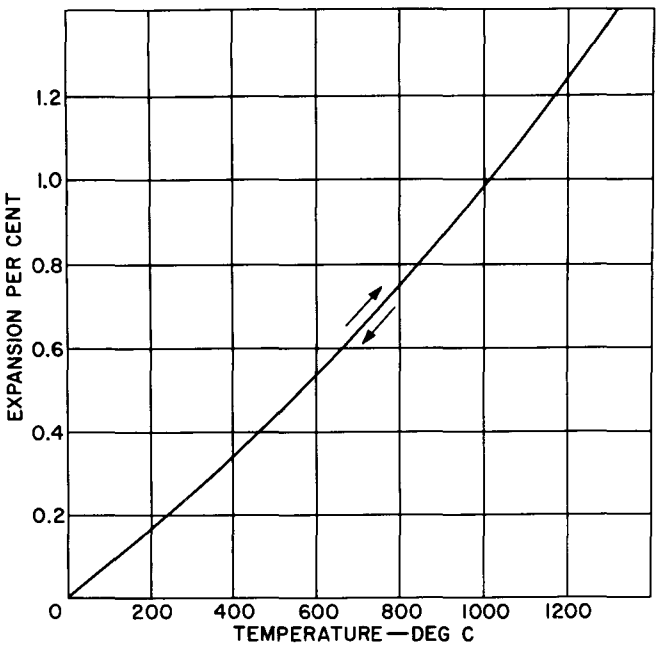


Figure 15. Expansion of Cubic (Stabilized) Zirconia as a Function of Temperature

Among the practical commercially available materials, forsterite compositions have the highest expansions (9.0 x 10^{-6}/deg C to 12.0 x 10^{-6}/deg C). Aluminas and steatites are intermediate (7.0 x 10^{-6}/deg C to 9.0 x 10^{-6}/ deg C). Zircon (4.0 x 10^{-6}/deg C to 5.0 x 10^{-6}/deg C) and cordierite (1.0 x 10^{-6}/deg C to 3.0 x 10^{-6}/deg C) are toward the low end.

A new group of ceramics—the lithium alumino silicates—exhibits unusual characteristics. Some have zero, or even negative, coefficients over an appreciable range of temperatures. Unfortunately, in their present state of development these materials cannot be easily fired to dense, dimensionally accurate shapes because of their extremely short firing range.

Table IV gives expansion data which are based on X-ray determinations of the change in lattice dimensions of crystalline materials as they are heated. This table is included because it gives some idea of the magnitude of difference that can be expected along crystallographic axes and over different temperature intervals. Fig. 16 shows graphically the range of values which are available in commercial ceramic materials compared with the individual, inherent value for elements and fixed compounds.

To use thermal-expansion data correctly, the designer must note the temperature ranges given. The most common method of presentation is to give one value covering room temperature to some elevated temperature (20 C to 800 C). Sometimes a series from room temperature to successively higher values is given (20 to 100, 20 to 200, 20 to 300, ··· , 20 to 900, 20 to 1000). A third method gives a series of successively higher narrow ranges (20 to 200, 200 to 400, 400 to

TABLE IV

X-Ray Thermal Expansion Data[14]

Name	Chemical Formula	Crystal Structure	Coefficient of Thermal Expansion (x 10^{-6}/deg C)		
			20 to 300 C	20 to 600 C	20 to 1200 C
Magnesia	MgO	Cubic	13.47	13.45	14.45
Corundum	Al_2O_3	Rhombohedral	9.39	9.48	9.50
Zinc Oxide	ZnO	Hexagonal—a axis	5.85	6.77	8.10
		Hexagonal—c axis	5.25	5.23	4.71
Rutile	TiO_2	Tetragonal—a axis	10.33	9.77	8.40
		Tetragonal—c axis	11.61	11.05	9.75
Spinel	$MgAl_2O_4$	Cubic	8.82	8.78	8.83
Spinel	$MgCr_2O_4$	Cubic	6.00	7.50	7.00
Spinel	$ZnAl_2O_4$	Cubic	8.09	8.31	8.95
Spinel	$ZnCr_2O_4$	Cubic	8.98	8.64	8.10
Beryllia	BeO	Hexagonal	8.22	8.44	9.02
Lime	CaO	Cubic	13.12	12.80	13.57
Strontia	SrO	Cubic	13.72	13.52	13.92

600, ··· , 800 to 1000). Although the values differ slightly from one system to another, the designer can be reasonably safe in making qualitative comparisons. For careful, precise work, such as the design of ceramic-metal seals, however, it is not possible to make valid comparisons unless all values are reported in identical fashion. In fact, large-scale plots showing the complete expansion curves of the exact ceramic and metal being considered are a necessity. Data of this sort should be continually reviewed and brought up to date as the manufacturer improves his materials or introduces new ones.

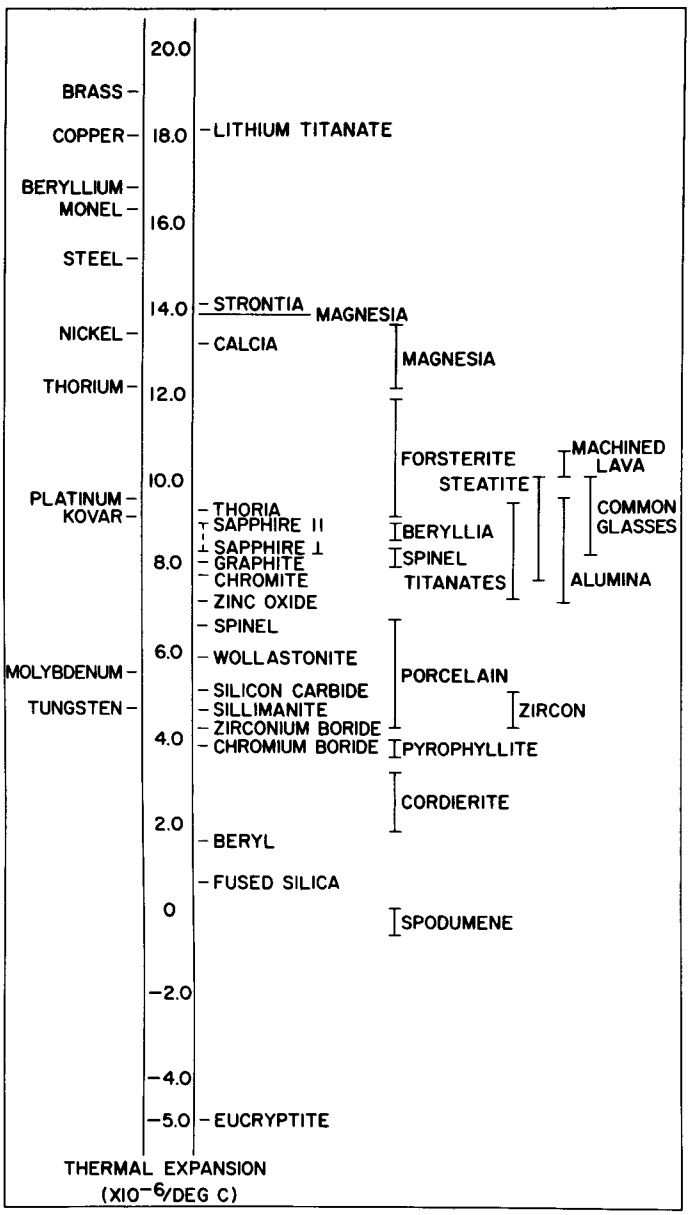


Figure 16. Coefficients of Thermal Expansion of Ceramics and Metals

Resistance to Thermal Shock

Resistance to thermal shock is the ability of a material to undergo very rapid changes in temperature without sustaining structural damage. This property has been evaluated in a number of ways, most of which are rather empirical. A method in wide use requires that the article be placed in a furnace at 100 C for a specified length of time. It is then removed and plunged into ice water. If no flaws are apparent, the article is put back into the furnace at 200 C. Repeated dips into ice water from successively higher temperatures are continued until the article cracks or shows signs of failure. Materials of the same shape and size are rated by the maximum temperature they will withstand.

In another method, a fixed temperature is used, and the cycles into ice water are continued until failure. The material which withstands the greatest number of cycles is rated best.

Both methods are time-consuming and reproducibility and sensitivity are poor. It is difficult to make valid distinctions between similar materials and more difficult to predict what will happen in actual service.

As a guide to thermal shock resistance, the following empirical expression developed by Winkleman and Schott has long been used. Recent work at Rutgers[9] confirms its general validity for rough approximations.

$$W = \frac{T}{EN}\left(\frac{K}{PC}\right)^{1/2}$$

where W = thermal shock resistance
T = tensile strength
E = modulus of elasticity
N = coefficient of thermal expansion
K = thermal conductivity
P = specific gravity
C = specific heat

Although Winkleman and Schott's equation places tensile strength and modulus of elasticity on a par with thermal expansion, practical experience has shown that thermal expansion is probably the most important single factor in thermal shock resistance. The higher the thermal expansion, the greater the possibility of shock failure. Materials like forsterite (12.0×10^{-6}/ deg C) are more susceptible than cordierite (2.0×10^{-6}/ deg C) or spodumene (0/ deg C) because of this.

If the thermal expansions of two materials are comparable, second-order effects like strength and thermal conductivity become relatively important. Beryllia, with its very high thermal conductivity, is more resistant to thermal shock than is alumina even though both have about the same strength and thermal expansion. It takes a big change in the second-order effects, however, to influence shock properties compared with the influence of expansion (Fig. 17).

The resistance of a particular device to thermal shock is also a function of the shape and mass of the parts involved. Because failure is the result of stresses which exceed the strength of the material, many shock problems can be overcome by design changes which minimize sharp edges, holes, or cross sections which tend to concentrate stresses.

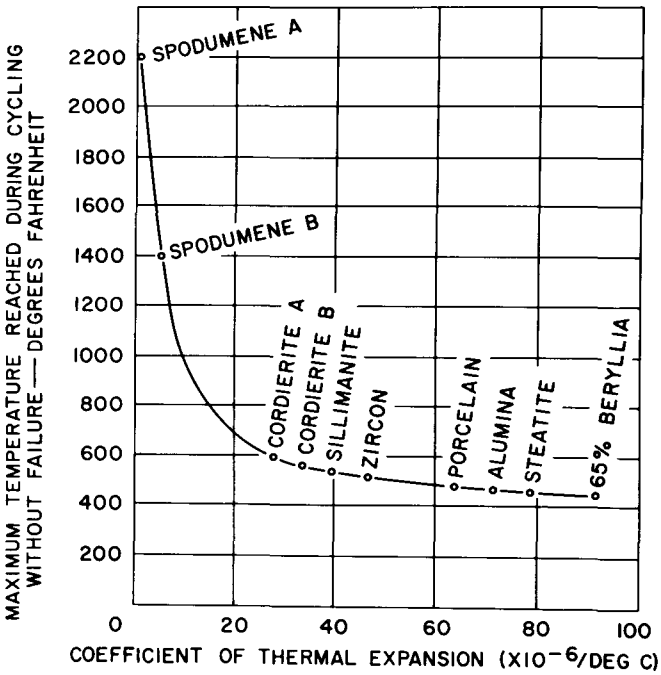


Figure 17. Thermal Shock as a Function of Coefficient of Thermal Expansion

ELECTRICAL PROPERTIES

Dielectric Constant and Loss Factor

Because there are certain types of tubes where high-frequency operation is of considerable importance, most ceramic data sheets give fairly complete information on loss factor, power factor, and dielectric constant as a function of frequency and temperature. At the present time, electrical properties need not concern the receiving-tube designer to too great a degree. The loss factor of most ceramics is sufficiently low to present no problems at the relatively low frequencies at which most receiving tubes operate. At high frequencies, however, the electrical properties are important. The dielectric constant and its influence on interelectrode capacity are, perhaps, more significant than loss factor in receiving-tube design.

Titania and the titanates have very high dielectric constants (K of approximately 100 to 20, 000). This, and other properties, limits their use as structural parts of a receiving tube. Among the ceramic materials more likely to be specified for a spacer, stem, or envelope, the range of dielectric constants is quite narrow as shown in Table V. At the high end are the zircons and aluminas (K = 9 to 10) and at the low end are the glasses and fused silica (K = 3 to 5).

Porosity exerts a strong influence on the dielectric constant since air and vacuum have a constant of 1.0. For any composition, the effect of porosity on the dielectric constant is practically linear. This phenomenon has been used to advantage in the fabrication of radomes. Foamed alumina and glass shapes have been made with constants approaching 1.0. Strength and thermal conductivity, however, decrease drastically.

TABLE V

Typical Dielectric Constants of Ceramic Materials

Material	Dielectric Constant (K)
Alumina	9-10
Zircon	8-10
Steatite	5-7
Porcelain	5-7
Forsterite	5-6
Cordierite	4-6
Pyrophyllite	4-5
Water	80
Titania	100
Titanates	1000-20, 000

In general, the dielectric constant increases with increasing temperature, but it is relatively unaffected, or decreases slightly, with increasing frequency.

When a perfect insulator, (i. e., a perfect capacitor) is subjected to an alternating emf, current will flow into the capacitor as the voltage is increasing and flow out as the voltage is decreasing. At the moment when the emf is a maximum, no current will be flowing, and when the emf is zero, the current will be at a maximum. Hence in a perfect capacitor, the current and voltage are 90 degrees out of phase. In an actual capacitor we do not have a perfect insulator and the current is not exactly 90 degrees out of phase with the impressed voltage. This difference between 90 degrees and the actual phase angle is called the loss angle. The sine of the loss angle (or cosine of the phase angle) is called the power factor. In ceramics, which are usually very good insulators, the power factor is synonymous with dissipation factor (tangent of the loss angle) because these values are practically the same when the phase angle is small. Hence, data sheets will usually give power factor as tangent of the loss angle or tan Δ.

The power factor, therefore, is a measure of the electrical insulating quality of a ceramic. Good insulators have extremely low power factors. Like the dielectric constant, the power factor increases with increasing temperature. As a function of frequency, however, it may decrease to a minimum and then increase.

The loss factor is the product of the dielectric constant and the power factor. This characteristic is extremely important in high-frequency applications. When the loss factor is too great, the ceramic begins to heat up. This effect soon leads to self-destruction since the loss factor increases rapidly with temperature. It is not uncommon to see portions of a contaminated alumina spacer melt (melting point 2050 C) as a result of losses within the part.

The ceramic suppliers are continually striving to develop materials with lower loss factors. Great strides have been made since World War II. Processing has been improved to eliminate metallic contamination, and purer raw materials have been specified which are free from such "lossy" impurities as alkali metals

(sodium and potassium) and iron-bearing minerals. The best insulator can be spoiled by careless handling or by poor tube making if it becomes coated with a film of dirt or vaporized metal, because these will heat up at high frequencies.

Specifications for insulators are listed in JAN I-10 military specifications. Six grades are given ranging from L-1 to L-6. L-6 grade is the best (some new compositions such as wollastonite are being called L-8 by extrapolation). Any material better than L-3 is called "low loss," and any value better than L-5 is called "ultra low loss."

Grade	Loss Factor	Grade	Loss Factor
L-1	0.150	L-4	0.016
L-2	0.070	L-5	0.008
L-3	0.035	L-6	0.004

Alumina, forsterite, and certain steatites are normally ultra-low-loss ceramics.

Values of loss factor for dense, vitreous pieces sometimes carry the designation "measured wet." Such a test is more rigorous and more sensitive than one "measured dry." The test pieces are soaked in water for 48 hours, after which the surface is dried, and the measurements are taken quickly. The slightest trace of water penetration into the piece will show up as a tremendous increase in the loss factor because water has a dielectric constant of 80. This imperviousness to moisture is an important criterion for materials which will be exposed to atmospheric conditions. Porous materials, of course, have to be measured "dry." Loss factors for various ceramic materials under different conditions of measurement are shown in Tables VI, VII, and VIII.

TABLE VI

Typical Loss Factors at One Megacycle

(measured "wet" at 25 C)

Material	Loss Factor
Forsterite	0.0014-0.002
Alumina	0.0014-0.011
Steatite	0.0025-0.0252
Zircon	0.006-0.0130
Alumina-mullite	0.010
Zircon-mullite	0.019
Lithia	0.0273
Cordierite	0.0297
Porcelain	0.040-0.058

Resistivity

The electrical resistivity of most ceramics is generally adequate for normal receiving-tube use. Resistivity is not an easy property to measure. Ceramic insulators have extremely high values of resistivity, therefore, precision instruments and careful techniques are required, especially when measurements are made at high temperatures.

TABLE VII

Typical Loss Factors at One Megacycle

(measured "dry" at 25 C)

Material	Loss Factor
Magnesium silicate	0.001
Lava	0.002
Alumina	0.003-0.011
Zircon	0.0041-0.0135
Magnesia	0.0048
Steatite	0.007-0.010
Cordierite	0.020-0.029
Aluminum silicate	0.046
Porcelain	0.054-0.060

TABLE VIII

Typical Loss Factors and Dielectric Constants at 10,000 Megacycles

Material	Loss Factor	Dielectric Constant (K)
Alumina	0.005-0.0218	8.0-9.6
Forsterite	0.0068	5.8
Steatite	0.008-0.030	5.3-5.8
Zircon	0.023	8.4

The normal practice is to rate materials in terms of the temperature at which a one-centimeter cube has a resistance of one megohm. The value is called the Te value. Representative values are shown in Fig. 18.

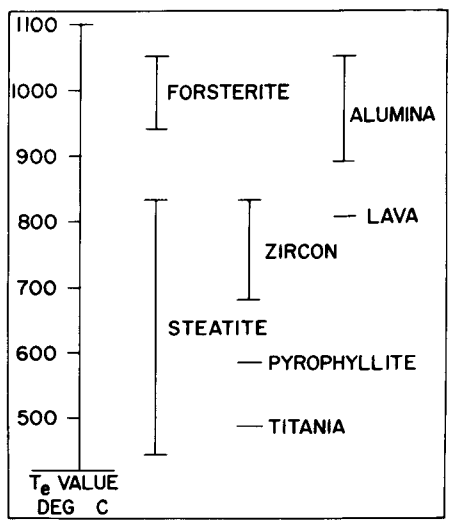


Figure 18. Te Value of Ceramics

Dielectric Strength

Dielectric strength refers to voltage breakdown and is given in volts per mil. It is important to know the sample thickness since thinner sections give higher values. Standard sample thickness is 0.250 inch.

In addition to thickness, two other factors also influence dielectric strength: temperature and the manner in which the voltage is applied. As temperature in-

creases, the dielectric strength drops considerably. Values are usually given at 25 C. Rapid application of power causes breakdown at lower values than does a slow, steady increase in voltage.

Table IX shows typical values of dielectric strength at room temperature (except where otherwise noted).

TABLE IX

Typical Dielectric Strengths at Room Temperature

Material	Dielectric Strength volts per mil
Alumina	200-300
Alumina (at 1100 C)	75
Zircon	250-300
Steatite	240
Forsterite	240
Electrical Porcelain	150-250
Cordierite	100-250
Lava	100
Pyrophyllite	80

REFERENCES

1. Kingery, W. D. (Ed.), Ceramic Fabrication Processes, Technology Press of MIT, Cambridge, 1958
2. Searle, A. B., Chemistry and Physics of Clays and Other Ceramic Materials, Interscience, New York, 1959
3. Newcomb, R., Ceramic Whitewares, Pitman, New York and Chicago, 1947
4. Hall, F. B. and H. Insley, "Phase Diagrams for Ceramists, "Journal American Ceramic Society , Nov., 1947, Part II; Dec., 1949, Part II ("Supplement No. 1")
5. Levin, E. M., H. F. McMurdie, and F. P. Hall, Phase Diagrams for Ceramists, American Ceramic Society, Columbus, 1956 (Part I); 1959 (Part II)
6. Insley, H. and V. D. Frechette, Microscopy of Ceramics, Academic Press, New York, 1955
7. McNamara, E., Ceramics, Vol. III: "Clay Products and Whitewares," Pennsylvania State College Press, State College, Pa., 1939
8. Kohl, W. H., Materials and Techniques for Electron Tubes, Reinhold, New York, 1960
9. Smoke, E. J. and J. H. Koenig, "Thermal Properties of Ceramics," Engineering Research Bulletin 40, Rutgers University, New Brunswick, N.J., 1958
10. Kingery, W. D., M. C. McQuarrie, et al, "Thermal Conductivity," Journal American Ceramic Society, Vol. 37, Part II, Feb. 1954
11. Norton, F. H., Refractories (3rd ed.), McGraw-Hill, New York, 1949
12. Ceramic Data Book (published yearly), Industrial Publications, Chicago
13. Wilson, H., Ceramics and Clay Technology (1st ed.), McGraw-Hill, New York, 1927
14. Beals, R. J. and R. L. Cook, Journal American Ceramic Society, Vol. 40, No. 8, p. 279

Ceramics and Ceramic-to-Metal Seals

M. Berg

Lancaster

CERAMICS FOR ELECTRONIC APPLICATIONS

Within the last few years, ceramics have assumed increasing importance in the electronics industry as a result of the development of widely varied applications such as ceramic tube envelopes, capacitors, memory devices, deflection cores, and radomes. In the past, the electronics engineer paid little attention to ceramics; he thought of them in much the same way as the ancient Greeks, to whom "Keramos" meant the finished product as well as the raw materials used in pottery. Nowadays, the science of ceramics is more narrowly defined as the high temperature chemistry of nonmetallic inorganic materials and includes noncrystalline materials such as glass.

Glass insulators were one of the few earliest uses of ceramics in the electrical industry; then, near the end of the last century, the development of vacuum-tight glass-to-metal seals made incandescent lamps a practical reality. The same type of enclosure, with some refinement in materials and processing techniques, was used by the radio industry for the first electron tubes.

In early receiving tubes, cathode heaters were inserted in ceramic sleeves to insulate the heater from the nickel cathode sleeve. This ceramic insulation is now provided more directly by coating the heater wire with a suspension of aluminum oxide sintered directly on the wire. As the electronics industry grew, increasing power requirements led to the use of ceramic spacers and support components inside the tube to provide more rigid spacing and higher power output. Recently developed ceramic envelopes for tubes have greatly improved power output, reliability, and compactness. In addition, generally improved performance under all operating conditions has resulted from the thermal, electrical, and mechanical properties of this type of envelope.

TYPES OF CERAMICS FOR ELECTRON-TUBE COMPONENTS

Ceramics are used extensively in electron tubes as electrical insulation spacers; the ceramic acts as a "solid vacuum", that is, it has the electrical properties of a vacuum and the mechanical properties of a solid. However, unlike a vacuum, the thermal conductivity of the ceramic selected for a specific application may be either high or low, depending on whether the insulated component is to run hot or be cooled by conduction. Although low dielectric loss and low dielectric constants together with high mechanical strength are generally required for most tube applications, for some special applications, high dielectric constants are desirable; however, the dielectric loss must always be as low as possible.

Insulators are graded by loss factor at a frequency of 1 megacycle as follows:

Grade	Loss Factor
L-1	0.150
L-2	0.070
L-3	0.035
L-4	0.016
L-5	0.008
L-6	0.004

Materials that meet these requirements are among the most recent developments in ceramic technology. Table I lists these "technical ceramics" and their typical properties. However, because each ceramic manufacturer has his own secret body composition and processes, the properties of commercial materials with the same name may vary from source to source.

The selection of ceramics for new devices or for substitution in old devices requires careful consideration to ensure maximum performance and economy. In selecting a ceramic material, the JAN 1-10 Joint Army-Navy Specifications on "Insulating Materials, Ceramic Radio" is an excellent guide. However, in the selection of such components, it is advisable to consult either a ceramic specialist or an experienced ceramic company. Brief descriptions of some of the various types of ceramics follow.

Alumina

Alumina (Al_2O_3) was probably one of the first ceramic materials used in electron tubes for heater coatings. However, only recently has alumina been extensively used for other insulators in tube construction. The alumina employed generally ranges from 90 to 100 per cent alumina, although some components range from 75 to 90 per cent. The high-alumina ceramics have excellent mechanical characteristics at room temperature and above and good dielectric-loss properties at

Table I

Characteristics of Technical Ceramics

Characteristics	Alumina	Steatite	Forsterite	Porcelain	Zircon Porcelain
Specific Gravity (g/cc)	3.1 - 3.9	2.5 - 2.7	2.7 - 2.9	2.3 - 2.5	3.5 - 3.8
Water absorption	0.0	0.0	0.0	0.0	0.0
Coefficient of linear thermal expansion per degree centigrade (20-700 degrees)	5.5-8.1 x 10^{-6}	8.6-10.5 x 10^{-6}	11 x 10^{-6}	5.0-6.8 x 10^{-6}	3.5-5.5 x 10^{-6}
Safe operating temperature (degrees centigrade)	1350-1500	1000-1100	1000-1100	1000	1000-1200
Thermal conductivity (cal/cm^2/cm/sec/deg C)	0.007-0.05	0.005-0.006	0.005-0.010	0.002-0.005	0.010-0.015
Tensile strength (psi)	8000-30,000	8000-10,000	8000-10,000	3000-8000	10,000-15,000
Flexural strength (psi)	20,000-70,000	16,000-24,000	18,000-20,000	9000-15,000	20,000-35,000
Modulus of elasticity (psi)	15-60 x 10^6	13-15 x 10^6	13-15 x 10^6	7-14 x 10^6	20-30 x 10^6
Thermal-shock resistance	Excellent	Moderate	Poor	Good	Good
Dielectric strength (v/mil: 1/4 in. thick specimen)	250-400	200-350	200-300	250-400	250-350
Resistivity (ohm/cm^3 at room temperature	10^{14} - 10^{15}	10^{13} - 10^{15}	10^{13} - 10^{15}	10^{12} - 10^{14}	10^{13} - 10^{15}
TE-value (deg C)	500 - 800	450 - 1000	above 1000	200 - 500	700 - 900
Power factor at 1 megacycle	0.0002-0.002	0.0008-0.0035	0.0003	0.006-0.010	0.0006-0.002
Dielectric constant	8 - 11	5.5 - 7.5	6.2	6.0 - 7.0	8 - 9
L-Grade (jan Spec. T-10)	L-2 - L-6	L-3 - L-5	L-6	L-2	L-4

low-through-high frequencies. The alumina ceramics are chiefly composed of alpha-type alumina crystals. Sintering is enhanced and grain growth is impeded by the addition of various foreign ions which, together with the alumina, form a glassy structure that binds the alumina crystal together.

For electronics applications, the tendency is to use the materials of high alumina content to obtain better mechanical and electrical characteristics; the difference in cost differential between a 90 and 98 per cent alumina is very slight. For ceramic-to-metal seals, the high alumina ceramics are definitely preferred because of their high mechanical strength and thermal conductivity. The electrical properties of these aluminas are in the range of grade L-4 to L-6.

Steatites - Clinoenstatite Bodies

Steatite was first used in Germany for low-frequency and low-voltage applications requiring close dimensional tolerances and good mechanical strength. It possesses better dielectric properties than the electrical porcelains previously employed.

Steatite is composed almost entirely of crypto-crystalline talc ($3MgO.4SiO_2.H_2O$). During firing, the talc is decomposed and clinoenstatite ($MgO.SiO_2$) and free silica are formed. Formerly, a small amount of clay and feldspar was added. The feldspar, which is admixed as flux, combines with the free silica and with the clay and its decomposition products to form a glassy

bond. However, because this glassy bond has an unfavorable influence on the electrical power factor, feldspar has largely been replaced by alkaline-earth carbonates, alumina, or a combination of these materials. These additives tend to eliminate the sodium ions and considerably improve the homogeneity of the crystal structure of the fired product; as a result, the dielectric properties are also improved. For extremely low-loss steatite, talc of very high purity and small crystal size is generally used as the raw material.

Some steatite bodies approach the grade L-6, but the more common commercially available steatites are graded L-4 or L-5.

Forsterite

Forsterite ceramics are mainly composed of the mineral forsterite ($2MgO.SiO_2$), which is formed commercially by adding a large percentage of magnesium carbonate or oxide to clinoenstatite bodies. Such compositions have particularly advantageous dielectric properties. The forsterites have a higher thermal expansion than the steatites, but are more frequency stable. However, as a result of higher thermal expansion, their resistance to thermal shock is lower than that of steatite.

For ceramic-to-metal seals, forsterite ceramics are quite suitable for compression seals in which the ceramic is maintained in compression by the metal component. Forsterite can be easily metalized by a number of techniques which are discussed in the latter part of this article. As shown in Table I, the material is not as strong as the high-alumina ceramics; however, its electrical loss properties and vacuum properties are at least equal to those of alumina.

Electrical Porcelain

For certain applications at low frequencies, porcelain which is made of china clay, ball clay, feldspar, and quartz, is still used in the electronic industry. Because of its plastic quality, clay can be formed into the most complicated shapes. It is fired at a temperature between 1300 and 1400 C. During the firing process, the feldspar dissolves the quartz and the cristobalite portion of the clay, (the clay is dissociated into the crystals mullite and cristobalite). The final structure on cooling is a conglomerate of mullite crystals, feldsparic glass and undissolved quartz particles. The presence of these glassy components, in particular, causes the high loss factor of the porcelains; they are classified in grades L-1 and L-2. However, these materials are less expensive than most other technical ceramics and, therefore, are frequently used in applications where their properties are suitable.

Zircon Porcelain

Compared to steatite and forsterite, zircon porcelains, introduced in 1943 as high-frequency insulators, have a wide firing range, higher mechanical strength, and higher thermal-shock resistance. The composition of commercial zircon porcelains varies; generally, the mineral zircon ($ZrO_2.SiO_2$) has been used as an admixture with other high-grade ceramic bodies to improve the body. The zircon porcelains are seldom used in electron-tube construction; however, these materials are replacing electrical porcelain in large electrical insulators. This type of ceramic, which can be fabricated to the density required to match the thermal expansion of molybdenum, is useful for ceramic-to-metal seal structures.

Beryllia

Beryllia (BeO) ceramics, among the newest materials in the electronics field, show great promise as electrical insulators and heat sinks. The material can be metalized and made into vacuum-tight enclosures for tube construction. Recently developed commercial beryllia materials have high strength (approximately 50 per cent that of the better alumina ceramics), but are quite expensive (four to ten times the cost of alumina). Because beryllia is highly toxic when finely divided or heated to 1100 C or above in a moist atmosphere, caution is necessary in handling the material.

The two outstanding properties of beryllia ceramics are its high thermal conductivity (approximately equal to that of aluminum) and its high electrical resistivity at high temperature. Beryllia ceramics are used for specialized applications such as heat sinks, heating cores, and microwave windows.

High-Dielectric Constant Ceramics

A new family of ceramics, based primarily on the mineral rutile (TiO_2) and the titanates of the alkaline-earth and rare-earth elements, has dielectric constants of 9000 and higher. The many different compositions developed have unusual electrical properties: some have positive temperature coefficients, others have negative coefficients, and some materials have piezoelectric characteristics.

Other Ceramics

Materials such as magnesia (MgO), fused silica (SiO_2), spinel ($MgO.Al_2O_3$), mullite ($3Al_2O_3.2SiO_2$), cordierite ($2MgO.2Al_2O_3.5SiO_2$), lithium alumino-silicates, boron nitride (BN), and the new family of Pyrocerams are also being investigated for use in tube construction.

Fused silica is used only in phototubes, which require high ultra-violet transmission. Although its electrical properties are very good, its poor mechanical strength and low thermal expansion make its application in power-tube construction almost impossible.

Boron nitride has been used for making prototype tube insulators, because its machining qualities are so superior; however, its outgassing properties ruin cathodes.

Magnesia has excellent electrical properties and thermal shock resistance, but is seldom used in tubes, although it is widely used as a spray for mica spacers.

Pyroceram, the trade name of a new class of ceramic materials developed by Corning Glass Works, is made by melting the chemical ingredients to form a glass.

The glass is then formed by standard glass-working techniques and devitrified by heat treatment to form fine crystallites. The material is stronger than glass and is pore-free.

Although many ceramic compositions must still be investigated for application in the field of electronics, the development of dense high-grade alumina has filled most of the designer's needs for the present and probably for the next few years. However, as power requirements, component size, strength, and other physical and electrical properties become more critical, the development of new materials will be required.

APPLICATIONS AND PHYSICAL PROPERTIES

Table II lists the more important ceramics and some of their typical applications in tubes. Fig. 1, which shows the effect of temperature on electrical resistance, indicates that most ceramics used in electron tubes have comparatively high electrical resistance at room temperature, but that the resistivity is greatly reduced at high temperatures.

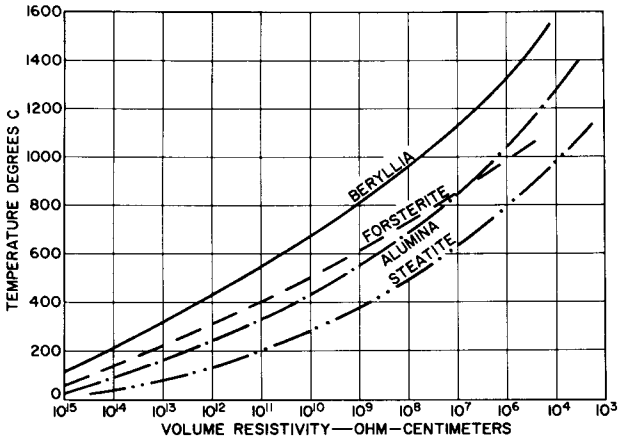


Figure 1. Variation of Volume Resistivity of Several Ceramics as a Function of Temperature

Linear expansion as a function of temperature, an important parameter for metal-to-ceramic seals and heat-shock resistance, is shown in Fig. 2 for various metals and ceramics.

Fig. 3 shows the loss factor as a function of frequency for several ceramics and glass; note how poorly glass compares to high Al_2O_3.

Design and Processing

The design of a ceramic component often governs the selection of the manufacturing process used in making the component. In the electronics industry, increasingly mechanized assembly has required new standards of dimensional uniformity and stability for all ceramic components. This demand, in turn, has stimulated the development of improved ceramic compositions and new fabricating techniques.

Size Tolerances. The ceramic components used in electron tubes should have wide dimensional tolerances. Most ceramic suppliers furnish components with plus or minus tolerances of either one per cent or 0.005 inch, whichever is larger. For closer tolerances, parts may have to be diamond-machined, specially sorted, or made on frequently renewed tooling. All of these additional procedures greatly increase the cost of the ceramics. In the designing and processing of ceramic tube components, the designer should work closely with personnel skilled in ceramics and with ceramic manufacturers.

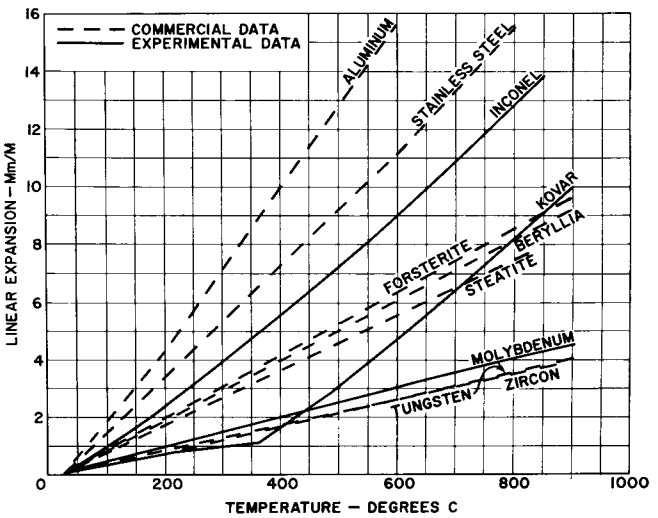


Figure 2. Thermal Expansion of Ceramics and Metals Used in Tube Construction

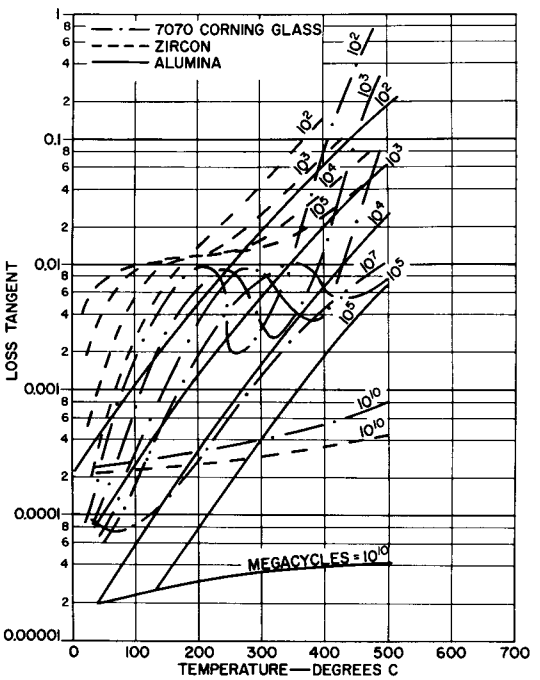


Figure 3. Loss Factor as a Function of Frequency and Temperature

Table II

Applications and Physical Properties of Technical Ceramics

Materials	Chief Chemical Composition	Application in Tubes	Mechanical Strength Flexure lbs./sq. in.	Thermal Shock Resistance	Max. Operating Temperature deg C	Thermal Conductivity g-cal/sec/ deg C/cm^2/cm
Alumina	Al_2O_3	Spacers, Coatings, Envelopes, Microwave Windows	50-75,000	Good	1650	0.04
Steatite	$MgO.SiO_2$	Spacers, Bases	10-20,000	Moderate	1000	0.006
Forsterite	$2MgO.SiO_2$	Spacers, Envelopes	10-25,000	Poor	1000	0.008
Electrical Porcelain	$3Al_2O_3.2SiO_2$	Bases	10-15,000	Fair	950	-
Zircon	ZrO_2SiO_2	Bases	10-20,000	Good	1100	0.012
Beryllia	BeO	Heat Sinks, Spacers, Microwave Windows	20-35,000	Excellent	1600	0.050
Titanates or Zirconates	$BaTiO_2$ $ZrTiO_2$	Mica Replacement	10-20,000	Fair	1000	0.012
Magnesia	MgO	Coatings	20-40,000	Good	1300	0.080
Fused Silica	SiO_2	Phototubes and Camera tube Faceplates	5-15,000	Excellent	1400	0.003
Boron Nitride	BN	Experimental Spacers	7-15,000	Excellent	1500	0.50
Pyroceram	Varies	Spacers, Envelopes	20-70,000	Good to Excellent	-	-

Porosity. The ceramics used for ceramic envelopes must be vacuum-tight, i.e., have practically zero porosity. For most ceramics, the theoretical density cannot be obtained, but a porosity of 3 to 5 per cent can be obtained with good processing techniques.

Porous materials are often used as interior tube spacers for their high dimensional accuracy, electrical properties, or economy.

In such cases, the following rule of thumb has been adopted to ensure proper vacuum exhausting of the porous structure. If the part is to be made porous, the porosity should be uniform throughout the structure and should range from 10 to 20 per cent. Porosity of over 20 per cent weakens the structure. When porosity is less than 10 per cent it is difficult to exhaust the porous structure. More than one per cent open pores makes the component difficult to degas. The recent trend is to use vitreous spacers throughout the tube structure because of their higher mechanical strength and reliability.

Strength. The mechanical strength of ceramics is a very important consideration in the design of tubes having ceramic-to-metal seals. High strength ceramics permit the design of more compact tubes having lower loss and higher power capabilities. Thinner cross sections result in higher thermal conductivity and low dielectric losses. For tubes having compression seals, such as the super power types, mechanical strength governs the selection of the ceramic; only ceramics of high strength can be used successfully for such applications.

In previous strength tests, test bars processed along with the components did not usually yield reliable data because they were not the same shape and size as the components. Very often ceramic manufacturers depend on the user to evaluate the product. More realistic testing procedures have been developed in which all tests are made on the actual component or on test samples machined from the component. Of all the insulating materials available, alumina ceramic is the strongest, beryllia next, and forsterite third.

Thermal-Shock Resistance. Resistance to thermal shock is quite important for ceramic components, because they are often subjected to severe thermal cycling during tube processing and life. The thermal shock resistance of the ceramic varies directly with the mechanical strength and thermal conductivity, and inversely with the thermal expansion and modulus of elasticity of the material. For most applications, the high strength of alumina makes it quite satisfactory for heat shock. One of the best materials is beryllia, because it has both high strength and high thermal conductivity.

Thermal Conductivity. Either high or low thermal conductivity may be required in various electronic components. For heater supports and insulations, low thermal conductivity is required; for most other applications, such as tube envelopes and insulating heat sinks, high thermal conductivity is desirable. Beryllia is by far the best material available; however, because of its high cost and toxic nature, its application has been very limited.

Electrical Properties. Every dielectric is characterized by its conductivity and dielectric constant. Conductivity, in general, varies with temperature and frequency. The component of conductivity which results from ionic conduction is little dependent on frequency; the component resulting from absorption losses is heavily dependent on both frequency and temperature. Both beryllia and alumina are excellent insulators at most frequencies employed in tubes.

The dielectric constant generally varies with temperature and frequency. Its component arising from normal polarization of the dielectric varies with temperature roughly in direct proportion to changes in specific gravity and decreases with frequency. The component arising from dielectric absorption varies directly with temperature and frequency and exhibits anomalous dispersion at regions of absorption maxima.

Both alumina and forsterite ceramics have low electrical losses at high-frequency and high-temperature applications. As the purity of the materials increases, the losses tend to become lower. Alkali impurities, such as sodium and potassium, are very detrimental; therefore, the fluxes for ceramics should be selected to maintain low-loss properties.

Fabrication Techniques

Dry Pressing. In the dry pressing process, the basic method used in forming most ceramic electronic components, dry granulated ceramic powders are pressed in metal dies under high pressure. This process is the most economical method of forming ceramics; for simple shapes, production rates as high as 2,000 pieces per minute are possible on some new presses. All compositions can be dry pressed; even the very abrasive pure oxide ceramics can be formed by automatic dry pressing when formulated with proper lubricants and binders. However, for large components other processes appear more feasible.

Extruding. Components having a uniform cross section, such as rods and tubes, are easily formed by the extrusion process. A plasticized ceramic mixture having a moisture content between 15 and 20 per cent is forced under pressure through extrusion dies. Excess air is removed from the ceramic body before extrusion for maximum density and packing. The extruded blanks are dried and can be machined before firing.

Isostatic Pressing. Because automatic dry pressing has limitations of size and shape, a new isostatic process has been developed for large or complicated components. In normal dry pressing, pressure can only be exerted in two directions through the movement of top and bottom punches. In the isostatic process, the ceramic powder is compressed in a flexible rubber mold, surrounded by a liquid, oil, or water. Pressure applied to a fluid is uniformly distributed throughout the medium and, therefore, it is uniformly applied to the compressible rubber mold. The rubber mold compresses the ceramic powder from all directions and forms a compound of uniform density. Metal inserts in the rubber mold form holes or threads.

Compression and Injection Molding. Recently developed methods for forming technical ceramics also include compression and injection molding processes. The principle and machinery are identical with those of plastic molding. A volume of 10 to 20 per cent of plastic molding powder is carefully blended with the ceramic powder. The mixture is heated to its molding temperature and compression or injection molded, depending on whether a thermosetting or thermoplastic molding compound is being used. The molded articles are cooled and hardened in the mold. They are then carefully fired to remove the organic molding compound without disrupting the remaining ceramic and without leaving traces of carbonaceous matter. The space originally occupied by the organic plastic is closed during the sintering process.

Hot Pressing. Relatively few, if any, products made by the hot pressing process are used in the electronic industry. However, at this time, hot pressing does offer one of the best methods of forming high-purity beryllium oxide ceramics. In this process, the ceramic powder is compressed in refractory dies and is simultaneously subjected to pressure and heat between 1400 and 2000 C. Metal molds soften or oxidize under these conditions; the only satisfactory mold material known is graphite. Unfortunately, because graphite is a rather soft material is can withstand only a few cycles.

CERAMIC-TO-METAL SEALS

During the war, the Telefunken Laboratories in Berlin developed ceramic-to-metal seals and vacuum-tight tube envelopes made of ceramic materials instead of glass. Although the quality of the steatite ceramic made the process unreliable, small radar tubes were manufactured on a large production scale with a very high number of rejections for seal leakage.

Allied Technical Teams realized the superior electrical characteristics of such ceramic tube designs and encouraged the United States Government to place contracts for the development of ceramic-to-metal seals,

ceramic materials, and ceramic tubes. Today, practically all new designs employ ceramics and ceramic-to-metal seals.

Some of the many advantages of the ceramic-to-metal seal structure include:

1. The outgassing temperature for most tubes is limited by the softening point or vapor pressure of the metal members or the brazing material; this temperature is usually much higher for ceramic tubes than for glass-construction types. As a result, the getter can often be eliminated because the metal member acts as a getter when the tube is outgassed at higher temperatures.

2. The more rugged and stable seals allow the ceramic tube to be operated at higher temperatures and, therefore, to dissipate more power than a glass tube of similar dimensions.

3. The ceramic structure lends itself more readily to close dimensional tolerances. Because ceramics permit very accurate spacing which is stable during the processing, tube characteristics can be more accurately controlled.

4. At high frequencies, ceramic structures have lower loss factors than the best glass structure and, therefore, for a given power, cause less dielectric heating.

5. The ceramic components lend themselves more readily to automatic production than glass components.

6. The high mechanical strength of ceramic structures permits their use in critical applications such as in missiles.

At present, the main disadvantages are the high cost of the ceramic materials and difficulties in processing the seals for good thermal-expansion match between the ceramic and sealing metal. These disadvantages are being rapidly overcome as new techniques for making strong seals, stronger ceramics, and lower cost ceramics are developed.

Applications in the electronics field range from the smallest devices such as tunnel diodes to such large devices as super power tubes and the C-Stellarator used to study hydrogen fusion. At this time, several manufacturers are actively investigating the feasibility of commercial production of receiving tubes having ceramic envelopes. Although ceramic envelopes have not yet been tried for kinescopes, phototubes, and camera tubes, X-ray tubes are available with ceramic envelopes.

PROCESSES USED TO JOIN METAL TO THE CERAMIC ENVELOPE

Metal is joined to the ceramic envelopes of electron devices by four general techniques: (1) sintered metal, (2) active alloy, (3) pressure, and (4) glass interlayer. Many variations in materials and processing are employed to obtain the most suitable end result for a particular structure.

The Sintered Metal Process

In general, this process employs finely divided powders of pure metals, mixtures of metals, or metal oxides. The powder, in a suitable suspension, is applied to the ceramic surface by printing, painting, or spraying and sintered to the ceramic by firing in a hydrogen atmosphere at 1250 to 1700 C. After sintering, the metallic surface may be plated with another metallic layer, such as copper or nickel, to promote wetting of the brazing materials; some brazing materials do not require plating.

The final step in the process is the joining of the metalized ceramic to the metal component by conventional methods of brazing in controlled-atmosphere furnaces. The exact mechanism by which the bond is formed probably varies for each combination of ceramic and metal powder. The metallic material appears to be bonded to the ceramic by one of the following processes: (1) solid-state volume diffusion of the metal into the lattice of the ceramic; (2) solid-state diffusion into the grain boundaries of the ceramic; or (3) liquid solution of the metal or metal oxide into the liquid or glassy phase of the ceramic. The author prefers to accept the first two mechanisms for ceramic materials of more than 95 per cent Al_2O_3 and the third mechanism for ceramics with less than 90 per cent Al_2O_3, forsterite, Steatite, zircon, and beryllia.

At RCA, a very extensive investigation of many combinations of metals and oxides for this process resulted in the development of a technique using pure molybdenum metal powder, which was found to be superior to all others evaluated. Such combinations as the well-known 80 per cent Mo - 20 per cent Mn, Fe-W, Mo-Fe_2O_3, Mo-Mo_2O_3, Mo_2O_3, Mo-CaO, Mo-Cu, Mo-Clay, Mo-SiO_2, and many others were evaluated on alumina ceramics varying from 92 to 99 per cent Al_2O_3. The highest strength obtained was with pure finely divided molybdenum powder on high-strength alumina made of 94 per cent Al_2O_3, 3 to 4 per cent MnO_2, and the remainder SiO_2. The flexure strength of the seal was between 60,000 and 70,000 pounds per square inch, which equals the strength of the solid ceramic tested in the same manner.

The same investigation also revealed that the important factors in making this type of seal are: (1) particle size of the metal powder; (2) metalizing temperature; (3) length of time at metalizing temperature; (4) atmosphere of the metalizing furnace; (5) thickness of metalizing coating; (6) plating; (7) brazing material; (8) brazing atmosphere; (9) brazing time; and (10) brazing temperature. Each of these ten variables is capable of destroying or weakening the seal; however, with proper control, excellent reproducibility and quality can be obtained.

One of the most important factors in controlling reproducibility was the development of a test for seal strength for evaluating the effects of the various parameters. The flexure test developed, which employs

an accurately machined bar to insure uniformity, is expected to become a standard ASTM test.

The following description of the various steps employed in the sintered metal process will help the reader to better understand the effects of the various parameters. First, a uniform coating of 0.001 to 0.003 inch of metalizing powder is applied to the ceramic surface as shown in Fig. 4A. The powder is then sintered in a controlled-atmosphere furnace so that the metal diffuses approximately 0.0006 inch into the ceramic and the powder sinters densely, as shown in Fig. 4B, to about one-half the prefired thickness. A plating approximately 0.0001 inch thick is applied for eutectic (72 per cent Ag - 28 per cent Cu) solder, as shown in Fig. 5A; no plating is required for most other materials such as copper, gold, or Nicoro. Finally, the components are furnace-brazed in a controlled atmosphere as shown in Fig. 5B. The final seal consists of: (1) a metal member; (2) an alloy braze layer containing the original brazing material, plating, and metalizing; (3) a thin metalized layer having pores filled with the braze alloy; (4) the metal-to-ceramic interface zone; and (5) the ceramic. For an ideal seal, the break or failure should occur at the highest stressed area, either in the ceramic or at the metal-to-ceramic interface. The seal configuration and the physical properties of the seal components (solder, metal, and ceramic) govern the location of the maximum stresses. In the case of poorly processed metal-ceramic seals, failure may occur at the following points: (1) at the ceramic-to-metal interface with no trace of ceramic breakage or metalizing, as shown in Fig. 6A, (2) in the sintered metalized layer as shown in Fig. 6B or (3) in the ceramic as shown in Fig. 6C.

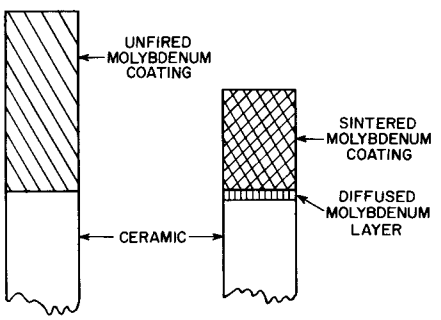


Figure 4. The Sintered-Metal Process: (A) before sintering; (B) after sintering

The particle size of the metalizing powder greatly influences the diffusibility of the metal into the ceramic and the density of the sintered layer. As shown in Fig. 7, the seal strength obtained with fine powders is more than twice that obtained with coarse powders. Failures such as shown in Fig. 6A result from coarse powders which form a weak layer and bond to the ceramic.

Metalizing temperature must be determined for each type of metalizing material and for each type of ceramic material. The variations in strength obtained at different temperature for a typical material, Almanox 4462, is shown in Fig. 8.

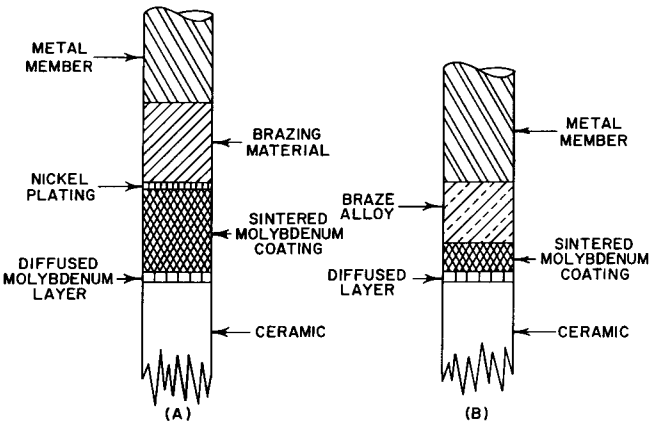


Figure 5. The Sintered-Metal Process: (A) before brazing; (B) after brazing

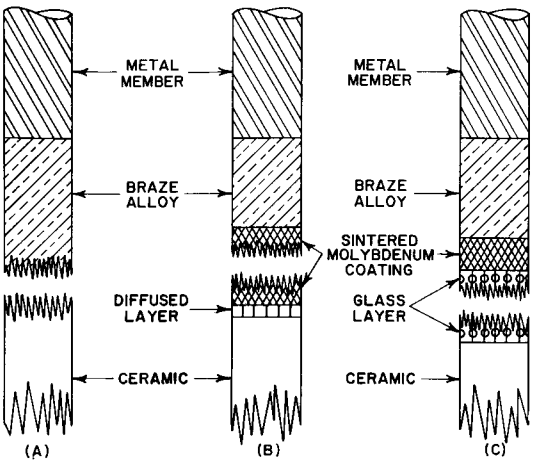


Figure 6. Failure Points in Ceramic-to-Metal Seals

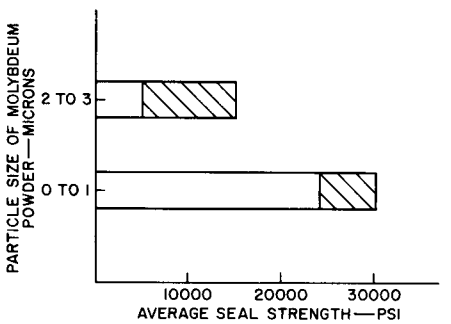


Figure 7. The Effect of Particle Size of Molybdenum Powder on Seal Strength

Failure caused by excessive temperature results from weakening of the ceramic by crystal growth or glass formation. If the temperature is too low, failure may result from a weak metalized layer and bond.

The length of time at metalizing temperature must be more than 5 minutes and less than 30 minutes at a

temperature in the range of 1450 to 1550 C. Less than 5 minutes may cause failure of the type shown in Fig. 6B, and excessive time may weaken the ceramic and distort its shape.

The atmosphere of the metalizing furnace is probably one of the most important of all the parameters. First, the furnace must be as gastight as possible to maintain the proper atmosphere. Various reducing atmospheres can be used - from 100 per cent hydrogen to mixtures containing a small percentage of hydrogen and an inert gas such as nitrogen, helium, or argon. Water vapor is added to enhance the sintering and diffusion of the metalizing material; the amount of water vapor required depends directly on the amount of hydrogen in the atmosphere. The higher per cent of hydrogen present, the larger amount of water vapor required. Free oxygen, in any amount, is harmful because it oxidizes the sintered powder. The effects of the atmosphere of the metalizing furnace on the strength of the seal (again with Almanox 4462) are shown in Fig. 9. Failures resulting from a furnace dry atmosphere would be of the type shown in Fig. 6A, and for an oxide condition of the type shown in Fig. 6B.

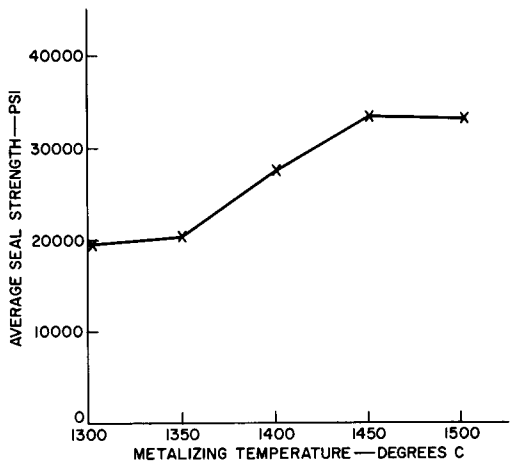


Figure 8. The Effect of Metalizing Temperature on Seal Strength

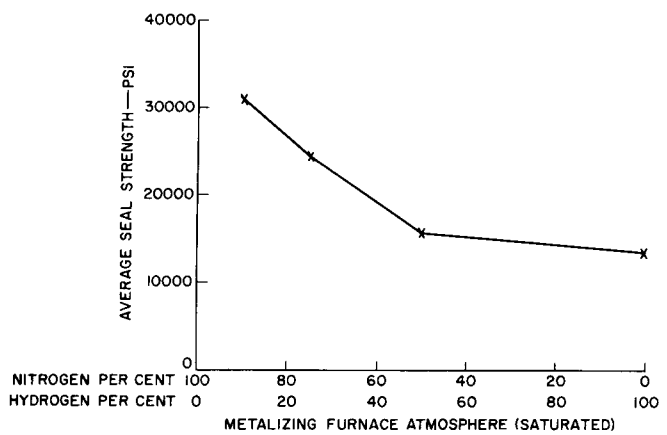


Figure 9. The Effect of Metalizing Furnace Atmosphere on Seal Strength

The thickness of metalizing coating is extremely critical for good seal quality. As shown in Fig. 10, the seal strength varies greatly with the thickness of coating. A coating which is too thin or too thick results in weak seals. A too-thick coating may result in a leaky and weak seal, because the fired metalizing layer is porous. The ideal thickness of an unfired coating is approximately 0.001 to 0.002 inch, and 0.0007 to 0.00015 inch for a fired coating. With the proper thickness and processing, the metalized layer is made nonporous during brazing by the brazing alloy which fills the pores; sufficient penetration is made into the ceramic composition to produce an excellent bond. The thickness of the metalizing coating can be controlled by processes such as printing or spraying.

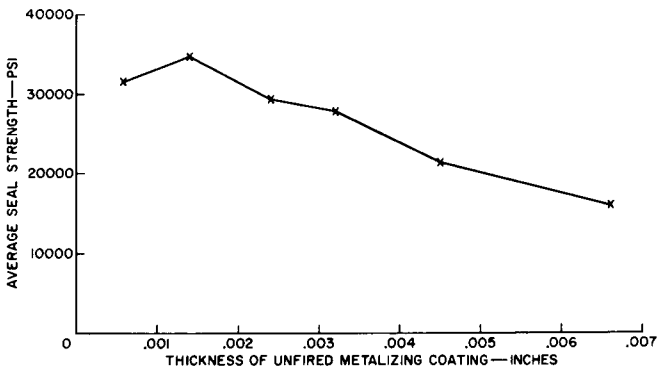


Figure 10. The Effect of Metalizing Coating Thickness on Seal Strength

Plating is quite important when brazing type solder is used, because the brazing material does wet the sintered metal sufficiently to ensure vacuum-tight, strong seals. The optimum plating thickness is between 0.0001 and 0.0003 inch. Poor plating may result in blistering during brazing and failure at the interface between the braze and the metalizing.

Brazing materials are quite numerous; a list of those commonly employed is shown in Table III. The ease of obtaining a good seal varies with the various brazing materials. Brazing-type solder, copper, the various eutectics, and pure materials are easy to use; materials without point melting properties are more difficult to use. Ductile solders having low melting points can be used most easily for making joints.

The brazing atmosphere must be free of contaminants that can react with either the braze or metalizing materials to form a poor wetting surface. A hydrogen atmosphere free of oxygen and low in water vapor is most suitable. Contaminated atmospheres may result in failures at the interfaces, such as shown in Fig. 6B, because of poor wetting.

The brazing time and temperature are closely related. In general, optimum seal strength is obtained by brazing for the shortest possible time at the lowest temperature. Either extreme of high temperature or long time may result in failures of the stripped seal type, such as shown in Fig. 6A. Too short a brazing time can

Table III

Brazing Materials

Brazing Material	Composition	Liquidus deg C	Solidus deg C	Vendor
Brazing Type (BT) Silver Solder	72% Ag - 28% Cu	779	779	Handy & Harman
"Incuro 60"	60% Au - Balance Cu and In	900	860	Western Gold & Platinum Co.
50 Gold - 50 Copper	50% Au - 50% Cu	950	930	Western Gold & Platinum Co.
Nicoro	82.5% Au - 17.5% Ni	950	950	Western Gold & Platinum Co.
35% Gold - 65% Copper	35% Au - 65% Cu	1005	970	Western Gold & Platinum Co.
Nicoro	35% Au - 62% Cu - 3% Ni	1020	990	Western Gold & Platinum Co.
10% Tin - 90% Copper	10% Sn - 90% Cu	1020	840	Made in Laboratory

result in a failure, as shown in Fig. 6B, where little alloying takes place with the solder and metalizing material.

The Active Alloy Process

The active-alloy process is preferred by some workers because only one firing operation in a high vacuum of the order of 10^{-5} torr, or in a protective atmosphere of helium, argon, or dry hydrogen, is required to make the complete seal. With the vacuum process, the tube can be exhausted during sealing and thus promote further savings. In this technique, an active metal such as titanium, zirconium, hafnium, or tantalum reacts with the ceramic body and forms a solid solution with the metal to establish an intimate bond with both members. The active metal powder or hydride may be printed or painted on the ceramic, or thin sheets of metal (shims) can be used. If one of the metal members is an active metal, a shim such as nickel can be used to form the alloy. The process has many variations depending on the results desired.

The important parameters of this process include: (1) the firing atmosphere, (2) the length of time at reaction temperature, (3) the amount and composition of active metal, (4) the composition of the ceramic, and (5) the composition of metal member.

Firing Atmosphere. Because the exclusion of oxygen from the seal area is one of the most important manufacturing conditions, active-alloy seals must be made in vacuum or in very pure hydrogen, argon, or helium. The release of undesirable gases such as oxygen, carbon monoxide, or carbon dioxide from the tube components must be prevented by careful outgassing of all metal parts prior to active-metal sealing.

The reaction temperature and the length of time at that temperature to a large degree control the reaction and formation of the solid solution of the alloy. If the time is too long, or the temperature is too high, the ceramic component may be weakened; normally, too-high temperatures increase the amount of metal taken in solution.

Variations in the amount and composition of the active metal have similar effects, and also determine the temperature at which the seal must be made. Table IV lists the range of temperatures that can be used with various active materials.

Although the use of hydrogen, helium, or argon atmospheres have been discussed for use with this process, as far as is known, the only reproducible results have been obtained with the vacuum process. The vacuum technique tends to limit this process for commercial applications. Although extensive strength testing has not been done for this process, the strength of the seal appears to be only 85 per cent of that obtainable with the sintered powder technique.

The Pressure Process

The pressure process is one of the simplest and oldest procedures used to adhere metal to ceramic. The two surfaces are forced against each other under pressure to obtain permanent coupling of the adjacent molecules of the two members.

Pressure seals to ceramic can be made with many metals, such as copper, indium, silver, gold, Kovar, and aluminum. In some cases, high temperatures are used to insure intimate bonding.

A variation of the pressure technique, in which metal members are used to maintain the pressure on the ceramic, has been successfully employed to make radial-compression seals for large power tubes. Fig. 11A shows an outside radial-compression seal arrangement before assembly. The outer diameter of the ceramic is larger than the inner diameter of the metal sleeve.

The seal is made by the application of an initial force,

Table IV

Active-Alloy Seals

Ceramic Material	Metal Part	Shim Material	Minimum Sealing Temperature deg C	Maximum Sealing Temperature deg C	Degassing Temperature deg C
Alumina, Forsterite	Copper	Titanium	875	910	1000
Alumina, Forsterite	Copper	Zirconium	885	910	1000
Alumina, Forsterite	Titanium	Nickel	955	1050	1200
Alumina, Zircon	Zirconium	Nickel	960	1050	1200
Alumina, Zircon	Zirconium	Iron	934	1015	1200

which may be as high as 20 tons, along the axis of the seal members. The wedging action creates a radial force F_R, as shown in Fig. 11B. The interference fit of the parts deforms both the metal and ceramic cylinders in the following proportion:

$$\frac{S_c}{S_m} = \frac{K_m}{K_c}\sqrt[4]{\frac{1 - r_c^2}{1 - r_m^2}}\left(\frac{R_c}{R_m}\right)^{1/2}\left(\frac{t_m}{t_c}\right)^{3/2}$$

where S is the unit strain in inches per inch, K is a constant (dependent upon locus or position of loading), r is Poisson's ratio, t is the wall thickness, and R is the mean radius.

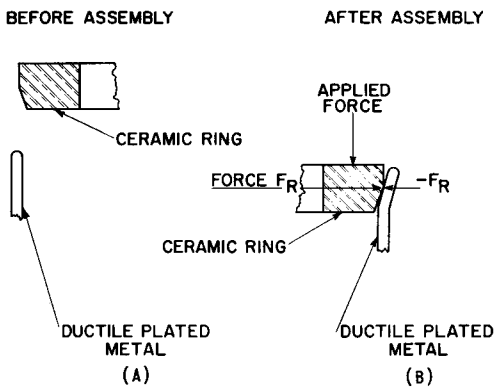


Figure 11. Radial-Compression Seal; (A) Before Assembly; (B) After Assembly

The subscripts c and m denote values for the ceramic and the metal, respectively. In most cases, the ceramic deformation is kept below 0.004 inch on the diameter. The hoop stress may be calculated as the quotient of the unit strain divided by the modulus of elasticity of the metal.

Fig. 12 shows a closeup view of the radial-compression seal. It can be seen that the force F_R is confined to a relatively narrow area of contact near the top of the taper. This high concentration of stress caused by the interference fit during sealing causes plastic deformation in the ductile plating, so that it flows around the ceramic to form a vacuum-tight seal.

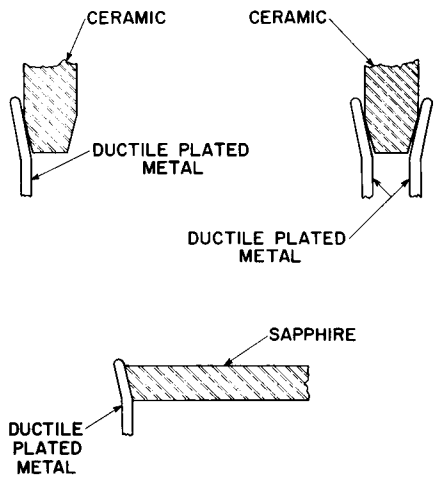


Figure 12. Three Types of Radial-Compression Seal: (A) Outside Seal; (B) Coaxial or Inside-Outside Seal; (C) Sapphire or Ceramic-Window Seal

The choice of materials for use in the radial-compression seal is quite important because the mechanical expansion and elasticity properties of these materials serve a definite function. Consider, as an example, the outside seal shown in Fig. 12A. Although the force maintained by the metal member on the ceramic is at a maximum at room temperature, it must exert sufficient force at maximum temperature to ensure a vacuum-tight seal.

The seals designed for the insulation of the C-Stellarator vessel are made from a high-alumina ceramic. The ceramic body has a flexural strength of about 60,000 pounds per square inch. However, high-strength metal suitable for a radial-compression seal has a much higher expansion coefficient than the ceramic. It can be shown that the expansion differential between the ceramic and the metal can be compensated for by the elasticity in the metal and ceramic.

Fig. 13 shows the stress-strain curve of a high-strength metal; this curve with some added information, illustrates the behavior of a radial-compression seal during bakeout. The force on the metal caused by the interference fit during the initial sealing process results in an elongation of the metal ring or sleeve to point A on the curve. This point is beyond the yield point of the metal. Although the stress in the metal sleeve at the sealing surface depends on the type of material selected, a typical value of 125,000 pounds per square inch is used in this case. The ultimate strength of the metal is in the order of 160,000 pounds per square inch. The total elongation in the metal sleeve at a stress of 125,000 pounds per square inch is 12 mils at room temperature. The total force at the seal surface of the metal-ceramic interface is about 1000 pounds per linear inch.

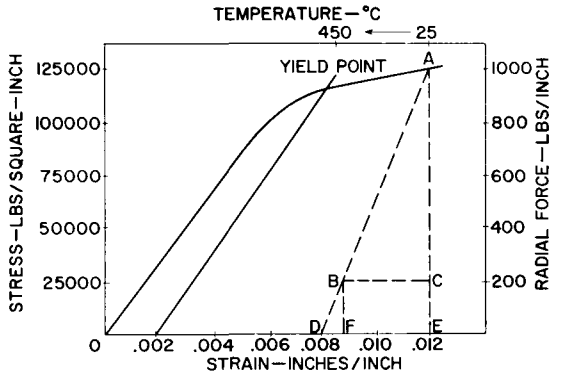


Figure 13. Stress-Strain Curve Showing the Behavior of a Radial-Compression Seal During Bakeout

If it is assumed that the metal has a greater expansion rate than the ceramic, line ABD, drawn parallel to the elastic portion of the curve, represents the slow release of stress as the temperature is increased. The total elastic strain storage in the metal sleeve at room temperature, represented by line DE, is about 4 mils per inch. Line BC is proportional to the product of the temperature range and the expansion differential. Therefore, if the two materials had identical expansion coefficients, the sealing force would remain constant. Any mismatch in expansion increases length BC and results in a reduction of force at the seal.

The stress in the metal sleeve cannot be permitted to reach zero. Some force must remain to maintain a vacuum-tight seal at the maximum temperature of operation. Because a portion FE of the total strain has been relieved by the differential expansion between the ceramic and the metal as the seal is raised in temperature, the remaining strain in the seal is now DF. This strain must produce sufficient force (for example, 200 lbs. per inch) to maintain the seal at this high temperature (in this case, 450 C).

The relationship used to determine the requirements at maximum bakeout temperature is as follows:

$$S_c + S_m - \Delta T \ \Delta C = S_b$$

where S_c is the unit elastic strain in the ceramic in inches per inch, S_m is the unit elastic strain in the metal in inches per inch, ΔT is the temperature difference in deg C, ΔC is the expansion coefficient difference per deg C, and S_b is the minimum value of strain in inches per inch.

This equation shows that the higher the strains and resultant stresses, the greater the permissible mismatch in the degree of expansion of the two materials. In these seals, only elastic strains are considered. The plastic strain is used only to increase the permissible tolerances of the parts, but otherwise serves no function during the temperature cycle.

The stress condition within the members requires careful consideration, especially because the longitudinal bending of the cylinder can result in a maximum stress an inch or more away from the metal-ceramic contact. These stresses can be minimized by the proper choice of parameters. The radial forces in the seal can be determined from the following expression:

$$V_o = KE(\Delta R)\left(\frac{t}{R}\right)^{3/2} \frac{1}{\sqrt[4]{1 - r^2}}$$

therefore, maximum bending stress X is given by:

$$X_b = \frac{K}{\sqrt{1 - r^2}} \cdot \frac{E \Delta R}{R}$$

where V_o is the radial force in pounds per inch, E is Young's modulus, t is the wall thickness, R is the mean radius, R is the change in radius, r is Poisson's ratio, K is a constant (dependent on locus or position of loading), and X_b is the maximum bending stress.

The Glass Interlayer Process

In the glass interlayer process, used for limited applications where electrical losses and mechanical strength is not important, a glass inserted between the ceramic and metal member can be used to make a seal. The technique is a combination of ceramic glazing and glass-to-metal sealing. The ceramic is first coated with glass frit, which is fired to produce a uniform glass coating on the ceramic. The metal component is prepared by the standard procedure for glass sealing and then furnace-sealed to the ceramic.

In one less frequently used process, a mixture of glass and a metal powder such as nickel, silver, or copper is fired onto the ceramic, which can be low-temperature soldered to a metal member.

SEAL DESIGN

There is practically no limitation to the type and size of seals that can be made with combinations of ceramic and metal; however, caution must be used in the design of each type to ensure reliability. The radial-compression design has already been discussed under pressure seals. For the sintered powder process and active-alloy process, design parameters are empirical at this time.

The types of seal most often used include: (1) the butt seal, (2) the flexible joint seal, (3) the outside diameter seal, and (4) the pin seal. The butt seal is one of the most severe from the viewpoint of strain; however, the seal has many advantages for electrical and tube ruggedness. The best butt seals are obtained with materials such as Kovar, copper-clad materials, or molybdenum, which keep stresses to a minimum. The use of optimum-strength ceramic and equivalent metalizing strength is desirable.

In the flexible seal, a normal thin metal area relieves the stresses on the ceramics, as shown in Fig. 14.

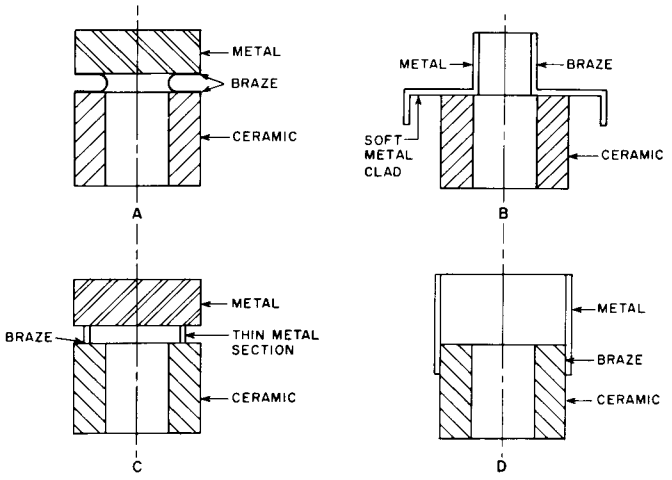


Figure 14. Various Types of Flexible Seals

The outside diameter seal is designed to maintain the ceramic in compression; however, tension failure can occur depending on the temperature cycling of the tube. This design has several limitations: the outer diameter of the ceramic must be accurately ground, sharp points are undesirable because the electrical field is highly concentrated, and the metalizing of the outer diameter is more difficult than the flats in commercial production.

Pin seals are of the more difficult types to make. In one frequently used design, which produces reliable seals, a Kovar washer is butt sealed to the ceramic with a pin solder to the Kovar or molybdenum washer, as shown in Fig. 15.

In another method, the inner diameter of the hole is metalized and the pin is brazed directly to the inner diameter with a ductile solder such as copper. With this design, it is essential that the expansion of the solder-pin composition is near that of the ceramic.

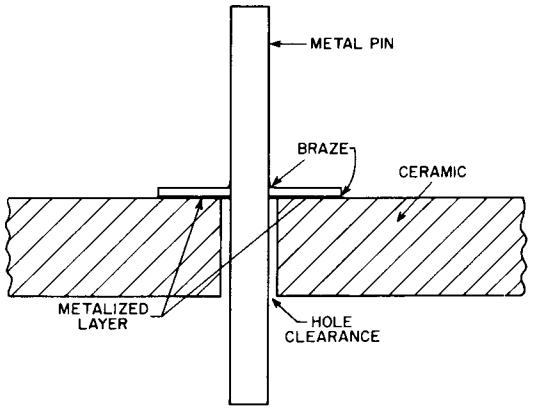


Figure 15. A Typical Pin Seal Design

BIBLIOGRAPHY

Ceramics for Electronic Applications

Code Y1

von Hippel, A. R., Dielectric Material and Applications, Chapter 4, The Technology Press of MIT and John Wiley & Sons, Inc., New York, 1954

von Hippel, A. R., "Tables of Dielectric Materials," Technical Report 119, Vol. V, Laboratory for Insulation Research, MIT, April 1957

Kingery, W. D., Ed., Ceramic Fabrication Processes, The Technology Press of MIT and John Wiley & Sons, Inc., New York, 1958

Kohl, W. H., Materials and Techniques for Electron Tubes, Chapter 2, Reinhold Publishing Co., New York, 1960

Norton, F. H., Ceramics, 3rd Edition, McGraw-Hill Book Co., New York, 1949

Norton, F. H., Elements of Ceramics, Addison-Wesley Press, Cambridge, Mass., 1952

Ceramic-to-Metal Seals

Kohl, W. H., Materials and Techniques for Electron Tubes, Chapter 14, Reinhold Publishing Co., New York, 1960

Martin, I. E., et al, Ceramic, Sapphire, and Glass Seals for the Model C-Stellarator, American Vacuum Society, October 1959

Code Y2

Code Y3

Stuart-Montieth, G., "Glass and Ceramic-to-Metal Seals," Brit. Communication and Electronics, pp. 439-443, 1961

Symposium on Ceramics and Ceramic-to-Metal Seals: April 21, 1953, Rutgers Univ. (Papers published in Ceramic Age, Feb.-Sept. 1954)

Van Houten, G. G., "A Survey of Ceramic-to-Metal Bonding," Jour. Amer. Cer. Soc., Vol. 38, No. 6, pp. 301-305, 1959

Design of Mica Spacers

M. Bondy and M. Jinetopulos

Harrison

Mica, because of its unique properties, is invaluable as an insulating spacer material:

1. Its cleavage characteristic makes it possible to obtain it in any desirable thickness.

2. Its high tensile strength makes it possible to punch small holes with close tolerances.

3. Its flexibility allows it to be force fitted or wedged into a glass bulb. Small spring points on the periphery of the shape can hold it rigidly in place, and can absorb vibration to minimize microphonic effects.

4. Its high electrical resistance minimizes electrical leakage between any elements held by it.

5. Its low power factor minimizes high-frequency losses in an electron tube.

6. Its chemical stability up to 600 C allows an electron tube to be processed and used at fairly high temperatures.

MINING AND PROCESSING OF MICA

Mica is a mineral, a chemical compound which occurs naturally as a product of inorganic processes.

Nine micas are known which differ in their chemical composition and properties, but all are grouped under the name mica because of their monoclinic crystal system. The chemical composition of mica is not pertinent to application in vacuum tubes, but water content and the ease with which water of constitution is lost when mica is heated determines the maximum temperature of its use. Muscovite, a potash mica, and phlogopite, a magnesium mica, are the only types of use in electron tubes; muscovite is the type now used by RCA. All micas possess the characteristics of perfect cleavage parallel to the basal plane, flexibility, elasticity, toughness, hardness, and transparency of thin plates.

Muscovite mica is usually ruby, green, or brown and is recovered in sheets from pegmatites, coarse-grained granite rock formations. The mining of mica is, at best, highly speculative. In the U.S. and Canada mica can be mined in a semimechanized manner. Full mechanization is impossible because the mineral bearing lode may be small and may run irregularly throughout a region. In addition, the use of tools damages the mica sheets. Thus, in America, most of the mica mined is scrap or small sheets. In India, manual labor is used exclusively. Care is taken to keep mica blocks intact and they are removed in the largest possible blocks. One of the largest mica crystals ever found yielded mica sheets of 30 x 24 inches. Normally, less than 1/5000 of the gross rock handled results in usable cut mica. The procedures used for obtaining usable mica are:

1. Crude mica is removed from the mine (usually a surface or open-cut mine). In crude form, mica is called "mine run" or "book."

2. Dirt and rock are removed from the crude mica in an operation called "cobbing." This is the first operation after the mica is mined.

3. Cobbed mica is split into sheets by an operation called rifting that consists of opening up crude mica books and splitting them into blocks from 0.007 to 0.030 inch thick.

4. The rifted mica is trimmed. The methods of trimming mica are:

a. Sickle-dress — the stone, foreign to the mica, is roughly removed with a sickle-like instrument.

b. Knife-dress — all cracks and other imperfections are cut out and bevelled edge is left all around.

c. Shear — similar to knife dressing except that there are no bevelled edges and the trim usually results in a rectangular shape.

d. Finger-dress (thumb trim) — loose edges are broken off by hand.

Usable mica is sold in the following forms:

1. Blocks — Knife-dressed mica 0.008 inch or more thick

2. Thins — Knife-dressed mica 0.001 to 0.008 inch thick

3. Splittings — Laminae split from blocks and thins

a. Loose splittings — heterogeneous, loose pack

b. Pan packed — loose but uniformly arranged and packed

c. Book-form — splittings arranged in books, as obtained from the same book mica

Mica is packed and sold according to size, thickness, and visual quality classification. The following factors determine the visual grade:

1. Clearness, uniformity
2. Air inclusions
3. Stains (smoky, sooty, black, red, green)
4. Waviness
5. Hardness (visually determined)
6. Stones and holes
7. Buckles, reeves, ridges, tears, cracks, etc.

Mica is a natural product for which there are no precise specifications for chemical composition or physical properties. Visual inspection is the only means of classification of quality. This method of classification depends upon the experience and judgment of the person doing the grading.

PROPERTIES OF MUSCOVITE

Some of the physical and chemical properties of muscovite are given in the following list.

Hardness (on Moh's scale)	2.8—3.2
Specific Gravity	2.6—3.2
Specific Heat	0.207
Optic Axial Angle—degrees	50—75
Volume Resistivity	2×10^{13}—1×10^{17} ohms per cubic centimeter
Surface Resistivity	10^{7}—10^{12} ohms
Dielectric Strength (volts per mil)	2800 in air for 1—3 mil thickness (less for thicker sheets)
Maximum Coefficient of Expansion, per degree C	0.000036
Modulus of Elasticity (pounds per square inch)	about 25×10^{6} (10 mil thickness)
Power Factor (1 megacycle)	0.0002
Dielectric Constant	6.5—8.7
Loss Factor (1 megacycle)	0.0013
Water of Constitution	4.5 per cent
Temperature at which Water of Constitution Is Driven Off	600—800 deg C
Maximum Temperature of Use	600 deg C
Resistance to Acid	Attacked Only by Hydrofluoric Acid

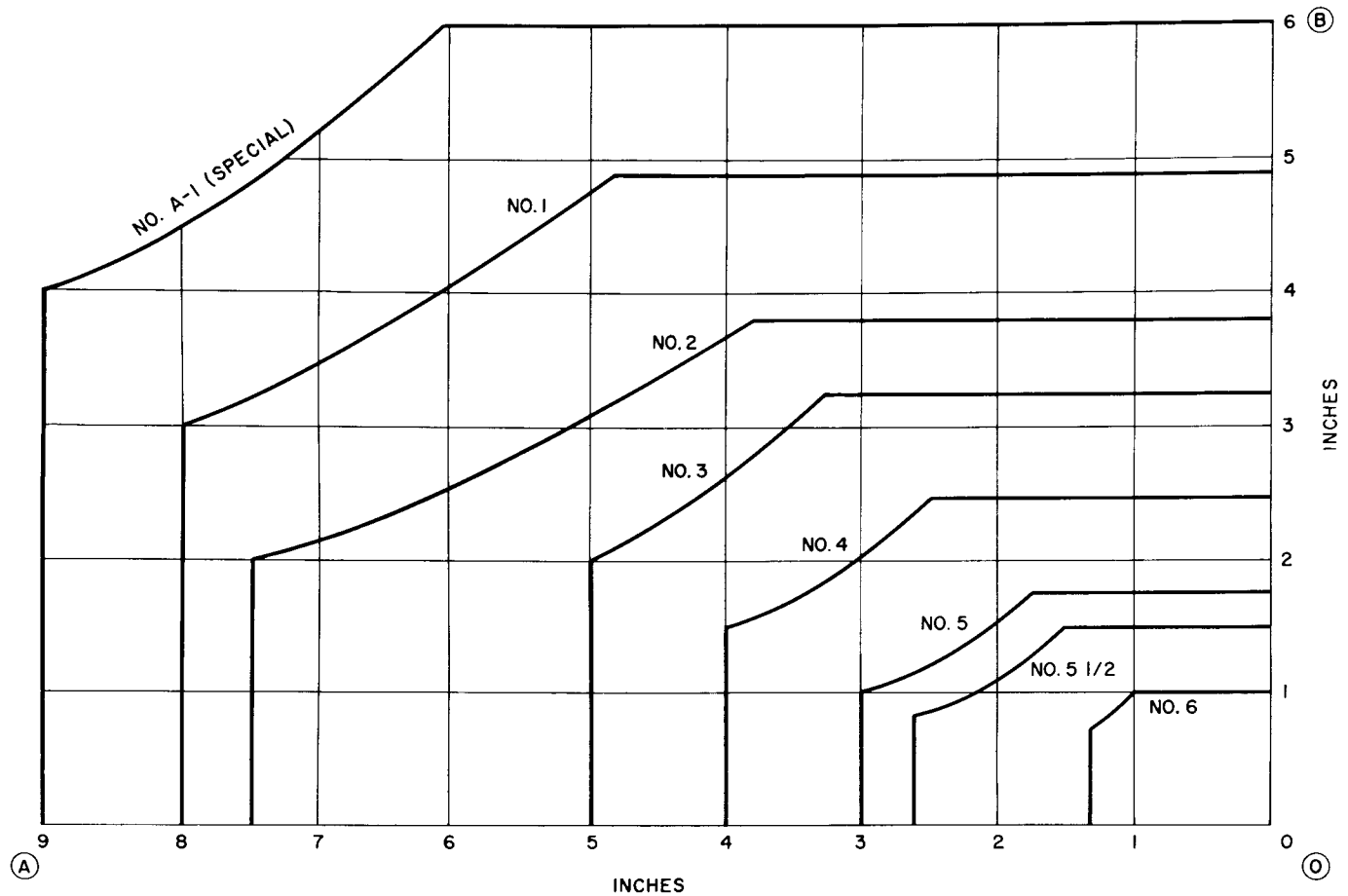


Figure 1. A.S.T.M. Chart for Grading Material Muscovite Block and Film Mica

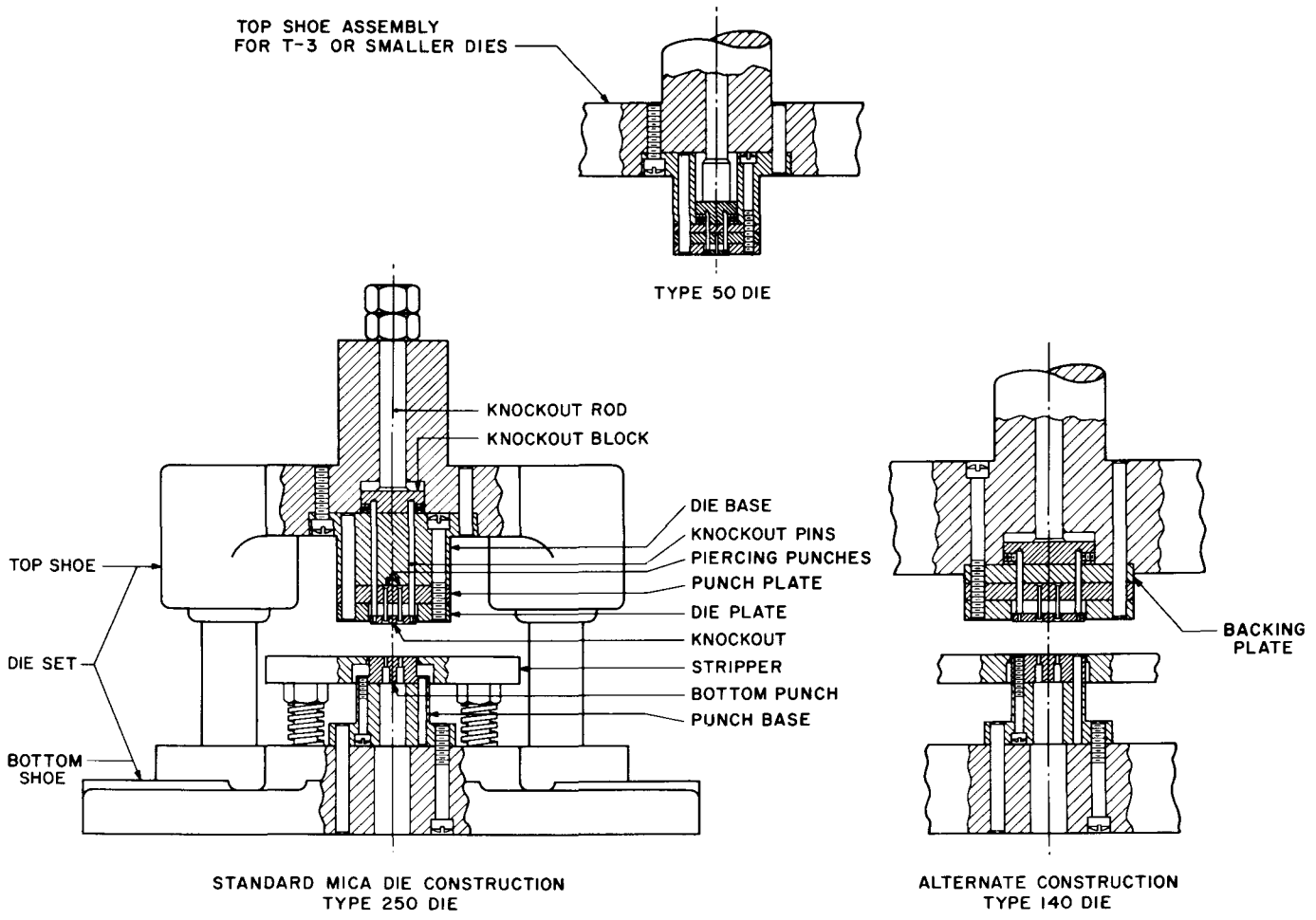


Figure 2. Typical Mica Die Construction. (Courtesy Schneider & Marquard, Inc.)

GRADING OF MICA

Mica is supplied in small, irregular pieces which must be individually hand loaded into the die. The pieces are graded by the American Society for Testing Materials[1] as shown on the chart of Fig. 1. Block mica and thins mica are graded for size as follows: The mica specimen to be graded is laid upon the chart so that it covers point O and has its maximum and minimum dimensions extending along lines OA and OB, respectively. The specimen is then shifted until the usable area completely covers the largest rectangle, determined by a diagonal extending from point O to or beyond a point on any of the curves. The number of the curve at the greatest distance from O cut by the diagonal of the rectangle designates the grade of the specimen.

TOOLING

Precision compound dies are required to produce clean-edged mica spacers free of burrs, delamination, and cracks. The die, even for the simplest design, requires meticulous die-makers and precision equipment in its manufacture. A different die is required not only for each change in pattern, but also for each change in thickness of the mica. Drawings of typical mica dies are shown in Fig. 2. The type number indicates the height, in mils, of the punches.

The typical die set for mica bridges is called the "cross-back leader-pin die set." The die is mounted in an inclined press and the mica bridges are transported through channel slides to sealed receptacles to protect the mica bridges from contamination. Accordion boots are used on die posts to protect the mica bridges from the lubricant used on the leader pins.

Fig. 3 shows a cross section of a compound die and illustrates the relative location of the parts. Notice that

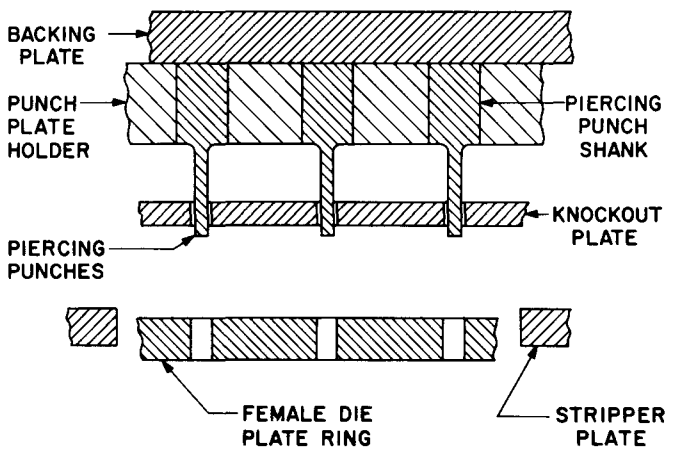


Figure 3. Cross Section of a Compound Die

the punch shanks occupy much more space than the punch. Within the contour of the mica bridge, space must be allowed for knock-out pins which strip the mica bridge from the punches. These knock-out pins, which are carried by another plate, pass through the backing plate and punch plate to operate the knock-out plate.

When mica is worked, highly abrasive dust particles are formed that could be injurious to the die. If allowed to accumulate, these particles would collect on all moving die parts to such an extent that, in short order, the parts would seize and rip themselves out of their holders. Air jets are built into the die set to disperse the dust particles.

Mica must be punched as free as possible from burrs, delamination, and cracks. To accomplish this, approximately 0.0003-inch side clearance is held between the piercing punches and the die openings. Provision is made in the knock-out plate for clearance of approximately 0.003 inch between the piercing punches and the knock-out plate. This space provides passage for the stream of air that blows the mica dust particles from between the piercing punches and knock-out plate.

SPACER DESIGN

The designer of mica spacers must bear in mind at all times the limitations of the material and the capabilities of the die. The fact that a design can be drawn on paper is no guarantee that it can be punched from mica.

Mica, because it is a natural product that is hand processed and hand selected, cannot be held to close thickness tolerances; tolerances as close as 0.001 inch require special selection and will result in marked increase in cost.

Distances between holes in the spacer are limited both by the strength of the mica and by the physical sizes of the piercing punches that can be made.

The minimum size of the holes that can be punched in a spacer depends upon the strength that a piercing punch must have to penetrate the required thickness of mica.

Very close tolerances are extremely difficult to hold and may require elaborate quality-control procedures and the use of standby dies. Close tolerances may, as a result, increase the cost of parts to four or five times as much as is usual.

Mica Thickness

RCA has standardized five ranges of thicknesses of mica. Table I gives these thicknesses in terms of bogey values and tolerances, and in terms of thickness ranges. Spacers must be designed to use one of the standard ranges. Spacers with diameters of 0.700 inch or less should have a maximum thickness of 0.012 inch (0.010-inch bogey). Spacers with diameters over 0.700 inch should use a maximum mica thickness of 0.015 inch (0.0125-inch bogey) for high-volume production; large spacers for low-volume production, for example spacers for large power tubes, may be as thick as 0.018 inch (0.016-inch bogey). In rare instances very large spacers with few holes have been stamped from mica as thick as 0.020 inch.

Table I

RCA Standard Mica Thickness

Bogey and Tolerances—Inches	Range—Inches
0.0065±0.0015	0.005—0.008
0.0075±0.0015	0.006—0.009
0.010±0.002	0.008—0.012
0.0125±0.0025	0.010—0.015
0.015±0.003	0.012—0.018

Bridge Distances

The need for the heavy shank on the piercing punch limits the minimum distance between holes. For punches of 0.050-inch diameter or less, the minimum bridge distance between holes is 0.015. (Bridge distance is defined as the minimum width of the mica between holes as shown in Fig. 4.)

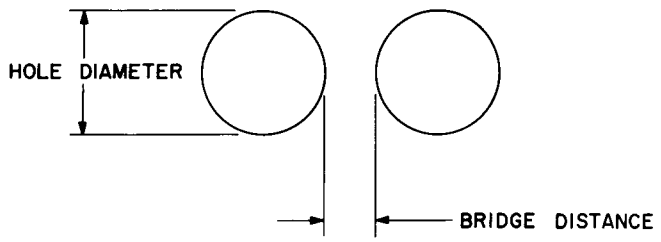


Figure 4. Bridge Between Two Round Holes

For hole sizes larger than 0.050 inch, the size of the shank of the punch approaches that of the punch itself and the bridge distance is limited by the strength of the mica rather than the punch design. Where little strength is required, as for leakage slots and degassing holes, the bridge distance may be less than 0.015 inch.

Minimum permissible bridge distance between any hole and the outer contour of the spacer is 0.030 inch. Designing for smaller bridge distances will result in breakage and delamination at the edge of the spacer.

Minimum Hole Dimensions

The minimum dimension of holes that can be punched through a given thickness of mica is fixed by the strength of the punch. Punches with a minimum dimension less than twice the thickness of the mica will be subject to breakage. Stated differently, the maximum thickness (the high end of the thickness range) of mica to be punched by a given die must not exceed one-half the minimum dimension of the smallest punch. Slots that have a minimum dimension less than half the mica thickness can be made if the punch is reinforced by ribs as illustrated by the "end-around" designs (Fig. 5).

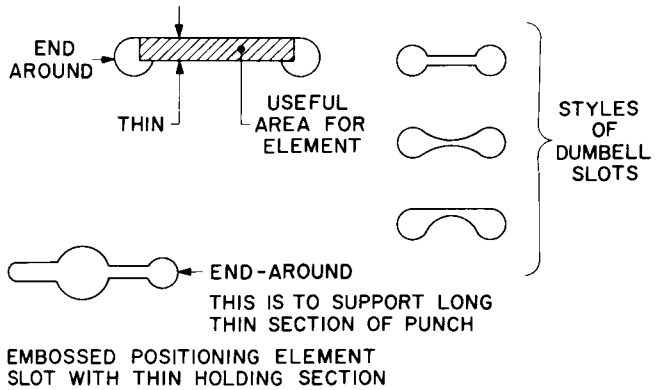

END AROUND
THIN
USEFUL AREA FOR ELEMENT
STYLES OF DUMBELL SLOTS
END-AROUND
THIS IS TO SUPPORT LONG THIN SECTION OF PUNCH
EMBOSSED POSITIONING ELEMENT SLOT WITH THIN HOLDING SECTION

Figure 5. End-Around Designs. These designs are used to strengthen the punches; thin sections and knife edges do not have long life

Outside Contour

When laying out a spacer design the first consideration is the contour of the periphery of the spacer. The periphery may have either a free fit or a force fit in the envelope (bulb) of the tube. If a free fit is desired, the contour should be as simple as possible—plain round or rectangular without sharp corners. If a force fit is required, the spacer is designed with points on its periphery as shown in Figs. 6 and 7. The diameter across the points is made slightly greater than the inside diameter of the envelope. When the bridge is forced into the envelope, the points bend slightly and give, in effect, a spring mounting. Because of the tolerances on the inside diameter of envelopes, the spacers are usually designed with two sets of points, one longer than the other. The shorter set of points is designated "A," and the longer set, "B." When the spacer is forced into an envelope with normal inside diameter (Fig. 6), the B points are in proper contact and the A points are free. When the spacer is forced into an envelope with a small inside diameter (Fig. 7), the B points will flex to the point of fracture and the A points will be in proper contact.

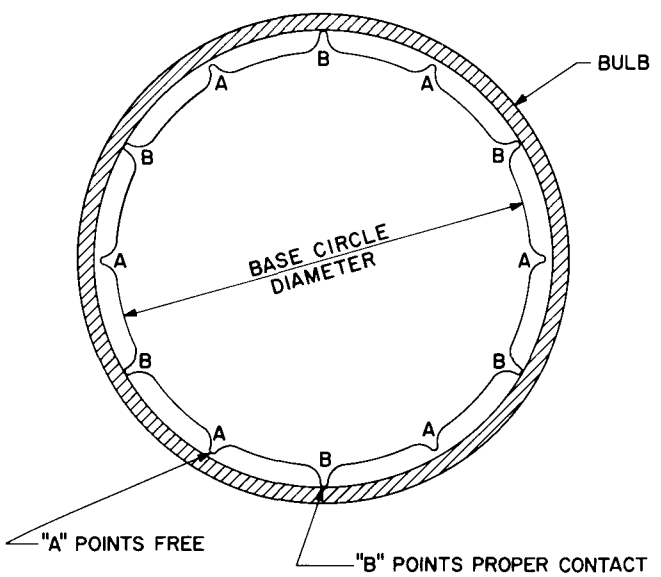


Figure 6. Cross Section of Bulb and Mica: normal inside diameter bulb with normal wall thickness

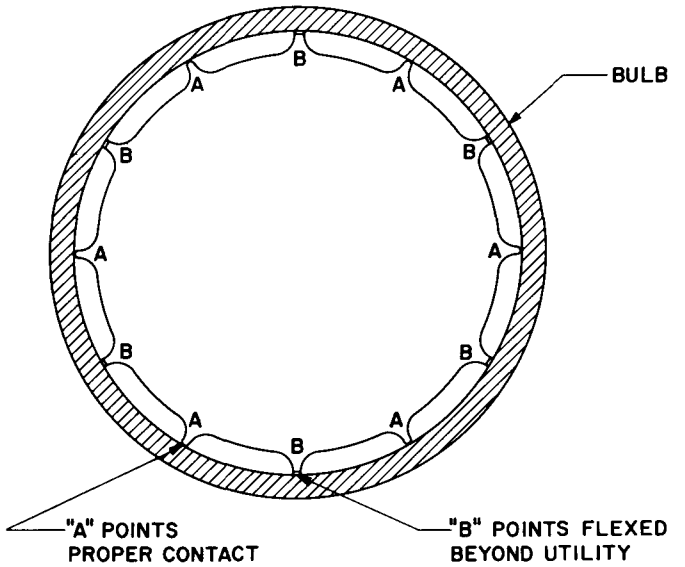


Figure 7. Cross Section of Bulb and Mica: small inside diameter bulb with maximum wall thickness

When extremely good force fits are required, the spacer may be made with three sets of points, or, alternatively, a single set of points may be used and the envelopes selectively fitted.

Design practice ordinarily requires a six-mil tolerance in diameter across the points; special designs may call for a four-mil diametral tolerance. The points should be designed with a large radius tangent to the end of the point and the base circle (Fig. 8). Lack of sufficient corner radius (Fig. 9) will cause excessive stress concentrations where the points meet the base circle and may result in fracture or delamination of the mica.

MICA COATINGS

During the operation of an electron tube, some parts, including the cathode, are at elevated temperatures.

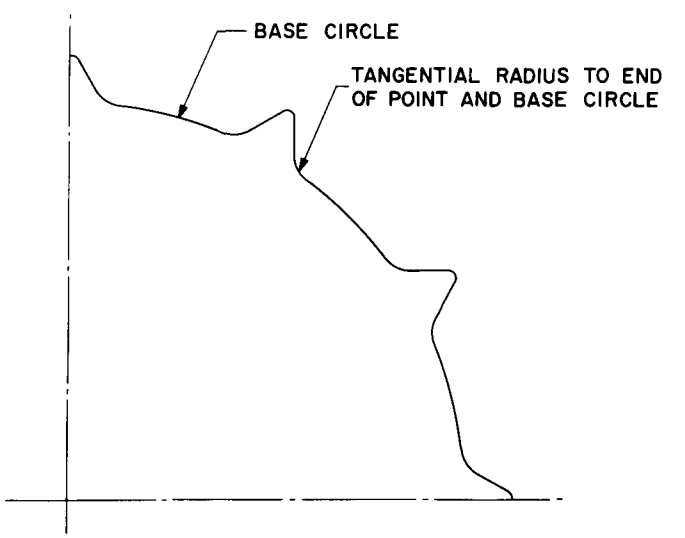


Figure 8. Preferred Contour of Mica Points

During the processing stages, the tube is operated at even higher temperatures to obtain cathode activation and proper degassing of the elements. At these elevated temperatures, some chemical elements, such as barium and magnesium, evaporate and deposit on the comparatively cool surface of the mica. When sufficient material has been deposited, a leakage path can be formed between the electrodes supported by the mica spacer. Factors such as the use of very pure cathode material may delay the formation of such leakage paths, but the possibility of their forming is still great.

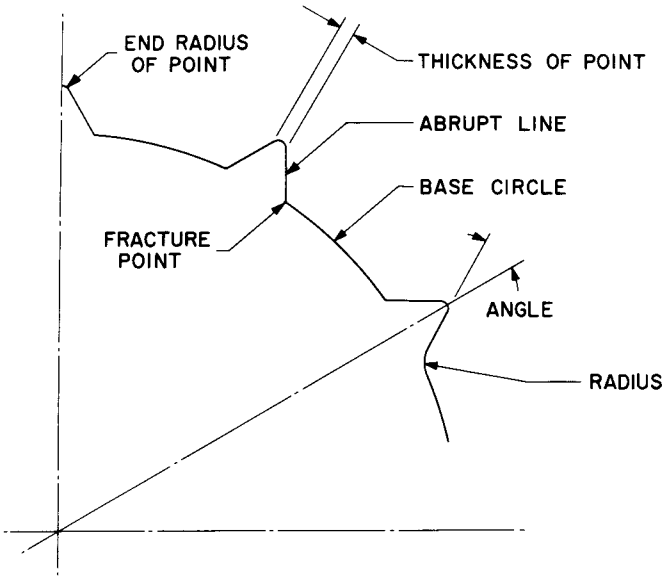


Figure 9. Improper Contour of Mica Points

The most commonly used methods of minimizing the effects of this evaporation film include the use of leakage slots, evaporation shields, double mica (Fig. 10), and fluffy mica coating. The best method is to coat the smooth surface of the mica with an insulating material which has a rough texture. This mica coating serves two purposes:

1. The effective surface distance between any two electrodes is lengthened. The series of hills and valleys created by a rough textured insulating coating means that a great deal more evaporated material must be deposited before a leakage path could develop.

2. Normally, the evaporated material comes from the cathode since the cathode is operated at approximately 1025 K, much higher than any other part except for the heater. (Vaporization from the heater must be driven out either end of the cathode and normally deposits on the bulb.) The deposition of material from the getter must always be controlled to minimize any settling on the mica. There is a certain amount of directionality associated with the "spray pattern" of the evaporant material. Because of this directionality, the rough surface casts "shadows" (Fig. 11) that prevent the deposited film from becoming continuous.

Tube types which must be very well controlled for

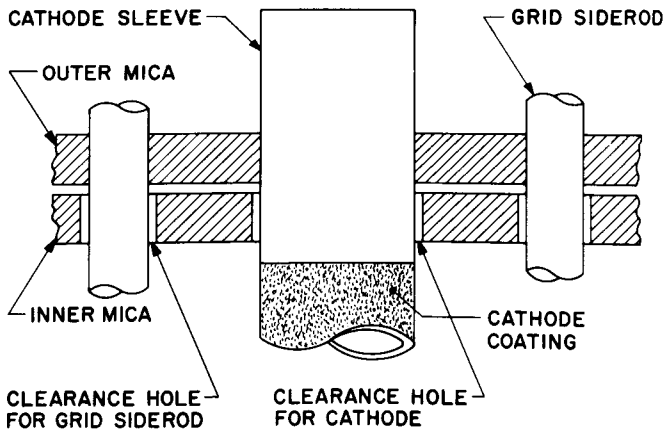


Figure 10. Cross Section of Double Mica Construction. Inner mica of "interelectrode-leakage" design

leakage should incorporate slots, evaporation shields, or double mica, in addition to the use of mica coating. Alumina, magnesia, powdered mica, and even chromium oxide are among the materials used to coat mica. The material is suspended in either water or a solution of silicates and sprayed on the mica. After it is coated, the mica is usually baked at a temperature of 400 C or more to remove water. Whatever the coating material, it must be a good insulator and must not volatilize, fuse, or decompose when heated. With its binder, it must be sprayable and have good adherence.

Tube types subject to high-voltage gradients or pulses require special mica coating material or treatment.

VOLTAGE GRADIENTS

Although some rough figures are available concerning the surface resistance of mica, care must be exercised to make any mica design as safe as possible from voltage breakdown[2]. Test results indicate that clean mica can withstand gradients as high as 125 volts per mil and mica with aluminum oxide mica coating has withstood gradients as high as 175 volts per mil before breakdown.

For practical reasons, gradients well below these values should be used when deposits of getter or other

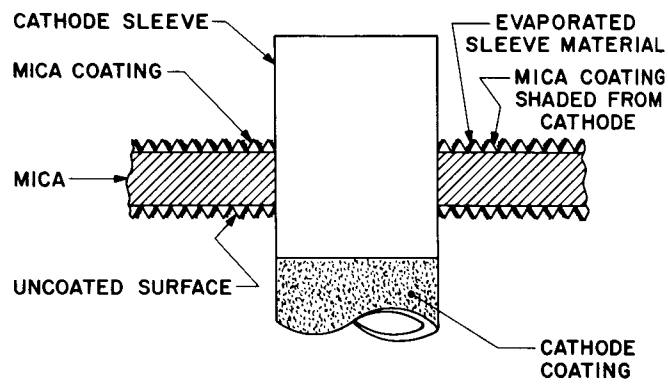


Figure 11. Cross Section of Mica in the Cathode Position Showing Sleeve Material Evaporated on Coated Mica Surface

evaporated metal may be present on the mica or when sharp metal parts are present in the mica.

Where gradients above 50 volts per mil seem to be needed, equipotential field plots[3] should be made of the mica and, where possible, leakage slots should be designed into the mica to make the best compromise between electrical safety and the mechanical mica considerations. Information on potential field plots will be found in the article by L. C. Scholz on Electric Field and Space Current, as well as in Ref. 3.

REFERENCES

1. "Specification for Natural Muscovite Mica Based on Visual Quality," D351, American Society for Testing Materials, Philadelphia, Pa.

2. Code K1

3. Sunshine Scientific Co., "Use and Theory of the Analog Field Plotter," (Instruction manual for apparatus manufactured by the Sunshine Co.)

Parts Cleaning and Processing

F. I. Scott

Harrison

IMPORTANCE AND SCOPE OF CLEANING

Cleaning is among the most fundamental aspects of successful tube making. The effect of the sturdiest, most reliable, and efficient tube materials and structures based on the most accurate theories can be nullified by inadequate cleaning. With a given set of materials and a given structure, improved cleaning results in improved tube performance and life expectancy.

The attainment of clean parts requires concern with practically all phases of parts manufacture and assembly. Some of the sources of contaminants in tubes are listed below. These sources range from the raw material itself to the temperatures attained during tube exhaust.

1. Materials
   a. Raw material itself
   b. Additives and coating materials
   c. Packing materials
2. Processing and Handling Conditions
   a. Lubricants and lubricating practices
   b. Storage environments
      1. Temperature and humidity
      2. Airborne chemical contaminants
   c. Processes
      1. Plating
      2. Washing and firing
      3. Miscellaneous treatments
   d. Handling and storage after cleaning
      1. Airborne contaminants (dust, lint, chemicals)
      2. Hand creams used by mounters
3. Sealing and Exhaust Conditions
   a. Sealing temperatures
   b. Vacuum-pump oil
   c. Track and rf-heating conditions

Cost is always important in any industrial undertaking. It will influence the cleanliness required, the type of cleaning used, and the location of cleaning operations in the over-all processing of components.

This article will discuss the contaminants encountered in tube-parts preparation, and their effect on tube performance. The major cleaning methods used in preparing tube parts will be considered in detail, after which the results of cleaning procedures on tube characteristics will be shown. Process and material controls related to attaining and maintaining clean parts will then be summarized.

Background

A perusal of RCA reports and recommendations back to 1939 reveals that there appears to have been no fundamental changes in cleaning methods or contamination problems. The major change in these areas has been in the degree of control exercised over manufacturing processes as well as over the cleaning processes. The use of cold carbon tetrachloride washing was eliminated in favor of hot trichlorethylene washing. Parts washing was shown to be superior to mount washing. Still, several tube types today beneficially use mount washing in addition to parts washing. Manufacturers have increased the stability of trichlorethylene, while the convenience, safety, and quality of the degreasing and other cleaning operations have improved with the introduction of better equipment. In general, modified approaches to the problems of cleaning are leading to better solutions of the problems.

A part of the reason for the slow progress and lack of radical change may lie in the wide variety of dirts which must be considered in cleaning problems. There is no universal cleaning method. The history of a material, as well as its future processing, must be considered in devising a suitable cleaning procedure.

General Considerations

Definition of Cleanliness. From a tube-maker's viewpoint, a clean part or surface is one carrying no material which will interfere with the desired operation of the tube, provided that sealing and exhausting processes are properly executed.

In a subsequent section, tests for measuring cleanliness are described. In a practical sense, a clean part may be defined as one free from materials detectable by these tests. For example, the most general specification for a clean part is that it be free of hydrophobic (water-repelling) organic contaminants and free of water-soluble contaminants.

Requirements for Cleaning. Two important conditions must be satisfied to obtain adequate cleaning: (1) the contaminant must be removed from the surface of the object; and (2) the contaminant must be removed from the vicinity of the object to prevent its redeposition. The type of contaminant will generally indicate the methods for its removal from the object. Removal from the vicinity may be accomplished by emulsification, chemical combination with the cleaning agent, or by repeated or continuous flushing with the cleaning media.

Classification of Contaminants

Many of the problems associated with cleaning stem from the great number of materials and substances which are contaminants for electron tubes. To permit a convenient approach to such problems, contaminants are usefully classified into five groups as shown in Table I. These classifications, while not mutual exclusive, or scientifically exact, do aid in systematizing the detection and removal of dirt.

Many contaminants cause visual changes in the cathode surface. These changes may assist in identifying the contaminant. Meltzer and Scott[1] have compiled and revised a summary of the effects and appearances of various contaminants.

A contaminant may not always be a substance "foreign" to the tube but may result from mislocation of an otherwise desirable material. Barium, for example, is a vital component of the cathode. But when it is deposited on grids, plates, or micas, it may so modify the tube characteristics as to cause malfunction.

Evolution of Gases from Parts

The kind, total quantity, and rate of evolution of gases from parts are important factors in the performance of tubes. Very small quantities (fraction of a micron) of reducing gases (e.g., CO, H_2) enhance emission. Similar quantities of oxidizing gases (e.g., H_2O, O_2, CO_2, SO_2) cause a decrease in emission by combining with free barium in the coating.

Gases ionize under some conditions of electron bombardment. Ions or atoms striking the oxide coating cause it to fuse where struck. The presence of ions between cathode and anode interferes with grid control of current flow and may significantly affect transconductance and cutoff. Depending on the kind of gas, electrode voltages, and electrode spacings, larger quantities of ionized gas will cause an arc discharge between anode and cathode which will destroy the cathode coating and sometimes the cathode base metal.

The gas associated with tube parts arises from three sources important to tube operation: (1) gases absorbed on the surfaces of materials; (2) gases dissolved in the materials; (3) decomposable compounds in or on the materials. The exhaust system removes those gases desorbed or decomposed by moderate heating under vacuum. The effectiveness of their removal accounts largely for the initial performance of otherwise satisfactory tubes. Dissolved gases and gases arising from diffusion-controlled decomposition of compounds affect the life performance of tubes. Not only the quantity of

Table I

Classification of Contaminants

Class of Materials	Examples	Possible Tube Defects Caused by Contaminant	Method of Contaminant Removal
Physically adherent	Dust, lint, particles	Shorts, leakage, noise	Flotation, ultrasonic agitation
Chemically adherent	Oxides, sulfides, phosphates	Low emission, secondary emission, gas	Chemical cleaning, firing in reducing atmosphere
Water soluble	Salt and similar compounds	Low emission, gas, bulb deposits, leakage	Pure-water rinsing, ultrasonic agitation
Solvent soluble (organic)	Oils, greases	"Black" cathodes, gas, low emission	Organic solvents, oxidizing treatments
Gaseous	Air, oxygen, carbon dioxide, water vapor	Low emission, early-hour transconductance slump, gas	Firing in reducing atmosphere, proper exhaust conditions

gas, but its rate of evolution is important. This rate will, in general, be determined by the diffusion rates of the gases or reducing agents in the metal.

Tests for Cleanliness

Many tests have been developed for estimating the cleanliness of parts. They range from qualitative tests for classes of substances to precise, quantitative tests for specific substances. For convenience in interpreting these measures of cleanliness, they are classified according to their degree of sensitivity in detecting small amounts, or small differences in amount, of contaminants, and according to their degree of reliability in predicting tube performance. The tests are summarized in Table II.

Table II

Tests for Cleanliness

Classification	Test	Contaminants Detected	Remarks
Qualitative	Ultraviolet radiation (3650A wave length)	Oils, fluorescent materials	Detects relatively large amounts of oil
	Microscopic examination (5X to 50X)	Lint, dust particles, corrosion	
	Hot plate	Nitrocellulose	For determining whether slurry-carbonized materials have been fired
	Peroxide-formic acid	Oxides	For determining whether gas-carbonized materials have been fired
	Residue	Oils	More sensitive than ultraviolet radiation
Quantitative	Chlorides	Chloride-containing salts	Should not exceed 20 micrograms per mount or 0.5 micrograms per milligram of cathode coating
	Vacuum extraction Vacuum fusion Mass spectroscopy	Gases and vapors	
	Miscellaneous wet chemical methods	Fluorides, sulfates	Fractions of a part per million detected
	Air oxidation (400-600 C)	Thin films of surface contaminants	Detects film 1-2 molecules thick
Semiquantitative	Emission spectroscopy	Metallic elements	Detects amounts down to hundredths of a per cent
	Water-break atomizer	Hydrophobic (water repelling) organics	Detects film approximately one molecule thick
	Selenious acid	Hydrophobic surface contaminants	Variable, similar to water-break test
	Conductivity	Water-soluble ionic materials	Detects as little as parts per million

Qualitative Tests. Qualitative tests detect only relatively large amounts of contaminants, and are primarily suited for quick factory checks on parts. Because large quantities of contaminant are required for detection, the reliability of these tests for predicting bad tubes is high. Failure to detect contaminants by these tests, however, does not insure good tubes.

The presence of oil is determined primarily by the use of ultraviolet radiation. Many oils fluoresce with a milky white color under exposure to ultraviolet radiation of 3650 angstroms wave length. Two of the limitations of this test are: (1) some oils do not fluoresce and (2) even fluorescent oils are not detectable below about 4×10^{-6} gm/cm^2 (Ref. 2).

Particulate matter such as lint or dust, and minute spots of corrosion or discoloration on parts may be determined by low-power (10x to 50x) microscopic examination. Detection of particulate matter may be done by examining the parts directly or by examining the surface of pure water while the part is being slowly immersed.

Quantitative Tests. Quantitative tests are characterized by high sensitivity and good reliability for predicting good or bad tubes.

Water or solvent residue tests predict tube performance fairly well and can be made quite accurate. Oils or other organic materials, in amounts below those responsive to ultraviolet radiation, are extracted by such solvents as ether, toluene, and trichlorethylene; similarly, deionized or distilled water may be used to extract and estimate the presence of water-soluble materials. In these tests, suspect parts or mounts are placed in the boiling solvent for five or ten minutes and then removed. The solvent is evaporated in a dish of known weight. Residues exceeding 1.0 milligrams per mount or 0.02 milligrams per part are generally primary or contributing causes of poor tube performance.

Extraction and measurement of chlorides provides important information regarding cleanliness. The method in use (photometric measurement of silver chloride precipitate) is very sensitive. The following limits may be taken as working maxima for chloride content: 20 micrograms per mount, and 3 micrograms per part, or 0.5 microgram per milligram of cathode coating[3].

Chlorides are particularly harmful to cathode emission because they react readily with barium and barium oxide.

Measurement and partial analysis of occluded, dissolved and combined gases are used to compare the purity of metallic and nonmetallic materials. Three basic methods of measuring gases are used in electron-tube making. They are vacuum extraction, vacuum fusion, and mass spectrography.

The vacuum-extraction test most nearly simulates the conditions experienced by the parts in tubes. It has wide use for practical comparisons of the gas levels of parts.

In the method, the part under test is heated in a vacuum in a system of known volume to a specified temperature and held for a specified time. The pressure rise resulting from the evolved gas is measured at room temperature. The gas content is reported as the product of the pressure and volume of the gas (microliters).

An extension of this test attempts to evaluate the gas released by a part during life operation of the tube. Following the test for initial gas content, the part is subjected to a time-temperature program which accelerates gas release. The program chosen is one that accelerates the diffusion of carbon in steel. In this manner the equivalent of 500 hours of diffusion is reached in a few minutes. Carbon is chosen because it is the most probable reducing agent to be found in metals and has the highest rate of diffusion.

The gas collected by the method may be partially analyzed by determining the portions condensible in liquid nitrogen, or in liquid nitrogen-acetone mixtures.

Vacuum fusion is used to determine the total quantity of dissolved and chemically combined gases in metals. In this test, the sample is placed in a carbon crucible and heated under vacuum to 1650 C to achieve complete fusion. The dissolved gases are expelled and the oxides reduced; the resulting gas is analyzed for hydrogen, oxygen, and less-active gases (predominantly nitrogen).

Mass spectrography provides both qualitative and quantitative information on the nature of gaseous materials. In this method, a gas sample is ionized by electric and magnetic fields to cause ions of varying charge-to-mass ratios to impinge on a detector plate. Qualitative analysis is based on the fact that the charge-to-mass ratio is a specific property of a substance, while quantitative measures are obtained from the magnitude of the current recorded at that ratio.

Analysis of fluorides, sulfates, and other negative ions are occasionally made in connection with tube problems. A variety of special tests may be used. Fluorides are a particularly potent poison for cathodes; they act similarly to chlorides to combine with barium and prevent emission.

Semiquantitative Tests. The following tests are quite responsive to very slight amounts of contaminants or very slight differences in amounts of contaminants. Because they detect contaminants in extremely small quantities, considerable effort is required to maintain the degree of cleanliness imposed by this test throughout parts preparation and assembly.

Oxidation of metallic parts at temperatures of 400 to 1000 C for 5 to 20 minutes is a sensitive test for contaminants. It is capable of detecting a layer of contaminant one to two molecules thick as evidenced by uneven oxidation or the appearance of interference rings on the oxide. Care must be taken that the test furnace and its atmosphere do not contain substances which will cause these defects.

Emission spectroscopy is used for qualitative analy-

sis of metallic substances and for quantitative estimates of some metals in concentrations below 1 per cent. Two major uses are: (1) inspection of parts and raw materials for trace impurities which may be detrimental to tube performance, and (2) analyses of bulb deposits to aid in tracing their source. The method can detect many metals down to about 0.01 per cent. Information from this test usually has a high degree of predictability for good or bad tube performance.

The water-break test consists simply of determining whether water spreads smoothly to a thin film on the surface of a part or forms distinct droplets on the surface. It is used to detect the presence of hydrophobic organic substances on a surface. It may be performed in two ways: the part may be immersed in pure water and withdrawn, or it may be sprayed with water from an atomizer. If the water forms a thin film, no hydrophobic substances are present. The immersion test is reported to be sensitive to 1 molecular layer and the atomizer test to 0.1 to 0.5 molecular layer of contaminant[4].

Changes in the conductivity (or resistivity) of distilled water can be used to detect and estimate the presence of ionizable water-soluble materials. Very pure water is only slightly conductive. If parts containing ionizable materials are immersed in pure water, the conductivity increase (or resistivity decrease) serves as a measure of cleanliness.

Methods of Cleaning

A number of cleaning techniques and procedures have been developed for removing industrial dirts. They are:

1. Heavy-duty methods
   a. Soak tank
   b. Mechanical or spray tank
   c. Emulsion
2. Medium and light-duty methods
   a. Solvent or vapor degreasing
   b. Water washing
3. High-temperature firing
   a. Reducing atmosphere
   b. Oxidizing atmosphere
4. Ultrasonic cleaning
5. Chemical cleaning
   a. Acid cleaning
   b. Brightetch
   c. Acetic acid
   d. Formic acid-hydrogen peroxide
   e. Electropolishing
   f. Alkaline cleaning, electrolytic
   g. Surface active agents (detergents, soaps)
6. Mechanical cleaning
   a. Sandblasting
   b. Burnishing

Soak-tank cleaning, mechanical-tank or spray cleaning, and emulsion cleaning are used for removing very heavy soils such as carbonaceous deposits and greases. Solvent degreasing, chemical and ultrasonic cleaning, and firing are used to remove the lighter soils and oxides such as are generally encountered in tube making. These methods, when properly used, leave minimal residues, a very important requirement from a tube making standpoint. The most widely used cleaning procedure for electron-tube parts consists of degreasing with trichlorethylene followed by washing in deionized (or distilled) water and firing in hydrogen at temperatures from 600 to 1000 C.

Trichlorethylene Degreasing. Trichlorethylene, marketed under several trade names such as Blacosolv (Blakeslee Co.), Triclene D (du Pont Co.), Nialk (Hooker Electrochemical Co.), and Permachlor NA (Detrex Co.), is a chlorinated hydrocarbon with the chemical formula HC_2Cl_3. It is an excellent solvent for oils, greases, and fats. The chlorine is covalently bound and does not readily hydrolyze so that there is less than 10 parts per million (ppm) of the free chloride ion. For metal degreasing, special inhibitors are added by the manufacturers to increase the solvents resistance to decomposition under a variety of conditions. Even so, the operation and maintenance of degreasers must be carefully controlled. Trichlorethylene is one-and-a-half times as heavy as water, boils at only 87 C, and requires approximately one-ninth of the heat needed to boil and vaporize it as is required by an equal quantity of water. It is inert to most industrial metals, nonflammable at ordinary temperatures, and has a high vapor density (4-1/2 times as heavy as air) which makes it desirable for vapor degreasing.

Among its limitations is the fact that soaps and soap-based lubricants are not generally soluble in it. (Soaps are not easily removed by any process.) Although trichlorethylene is less toxic than many other chlorinated solvents, it does have appreciable anesthetic and toxic effect, and requires adequate precautions in use. The recommended maximum allowable atmospheric concentration is 200 ppm for an eight-hour day. Periodic analyses are made to insure that this limit is not exceeded. The solvent can be decomposed by heat greater than about 120 C, ultraviolet radiation, acids, and alkalies. Excessive moisture increases vapor losses from the degreaser and aids the breakdown of the solvent to form free chlorides. Trichlorethylene degreasing does not leave a water-break-free surface.

The degreasing process employs a tank consisting of from two to five compartments arranged to provide overflow successively from the highest to the lowest compartment. From the lowest compartment, the dirty solvent is pumped to a still, purified by distillation and returned to the highest compartment. Steam coils heat the solvent; water-cooled coils around the tank walls above the boiling solvent minimize vapor losses. Baskets of parts are immersed in one of the lower compartments and then are immersed successively in higher compartments for a specified time (usually three to five minutes each). After the parts are drained on a rack at or above the vapor level, they are ready for further processing.

The degreasing operation is controlled by twice-a-week checks on the oil content of the solvent in the cleanest (highest) compartment. Average oil content of this compartment is 2.5 mg/100 cc (17 ppm). Biweekly, the acidity or alkalinity of the solvent is checked by measuring the pH of a water extract. The pH should not be below 6.5; normal operating level is 8.0 to 9.0. The degreaser and still are cleaned monthly, in addition, the tanks are boiled out at the same time with soda-ash solution. Depending on work load, the still and the lowest degreaser compartment may be cleaned weekly.

Water Washing. Water washing removes water-soluble materials which generally are not soluble in trichlorethylene. Pure water for this operation is obtained by deionization, a method for removing ionized substances from water. It differs from distillation in that it separates ionized materials by adsorption on synthetic resin rather than by vapor-pressure differences between water and the ionized material. Deionization is cheaper than distillation and can produce water having a lower solids content.

Deionization is accomplished by passing tap water through two kinds of synthetic resins. Either two-bed or single-bed systems may be used; higher-quality water is obtained by the latter method. One resin (usually a polystyrene-sulfonic acid type) adsorbs cations such as calcium and magnesium, and releases an equivalent number of hydrogen (H^+) ions. The second resin (often a polystyrene quaternary amine) adsorbs anions such as chloride and sulfate, and releases an equivalent number of hydroxyl (OH^-) ions. The hydrogen and hydroxyl ions, of course, form water. Quality of water produced by deionizers is rated as follows:

1. Two-bed deionizers containing low-basic anion resin; 3 to 5 ppm (resistance of water 100,000-170,000 ohms/cc)
2. Two-bed deionizers containing high-basic anion resin; less than 3 ppm (resistance greater than 170,000 ohms/cc)
3. Single-bed deionizers; less than 1 ppm (resistances of 1 to 20 megohms/cc)

Water washing normally follows degreasing of tube parts. Two overflowing compartments are used with the second (dirty) compartment overflowing to drain. Morpholine, an organic amine is added drop-by-drop to the water to keep it alkaline and prevent rusting of iron and steel parts. Following water washing, most parts are dipped in methanol to aid drying, which is done in a hot-air oven. Slurry (S95) carbonized parts are not methanol-dipped to avoid attack of the nitrocellulose binder.

A minimum resistivity of 250,000 ohms/cc is the specified control for inlet water to the wash tanks. The tanks are cleaned at the monthly degreaser cleaning.

Firing. Firing in hydrogen or other reducing atmosphere is the most widely used method for reducing metallic oxides. It does not effectively remove organic films of high molecular weight[4].

High firing temperatures and long times are desirable. Possible distortion, excessive strength loss, or vaporization of some metals, however, limits the permissible firing conditions. For these reasons a variety of firing conditions are employed. Generally, cathodes are fired for 5 to 10 minutes at 600 C; plates and other parts are fired for 7 to 10 minutes at 800 C. Aluminum-clad parts usually are not fired above 600 C to prevent alotization (formation of iron-aluminum alloy which causes welding difficulty). Other atmospheres used for firing parts include partially burned city gas, forming gas (mixtures of hydrogen and nitrogen), methane, and vacuum. Parts should be mounted and exhausted as soon as possible after firing.

The rapid reduction of oxides requires the maintenance of a hydrogen atmosphere having a low dewpoint. (Dewpoint is the temperature at which a gas is saturated with water vapor and, consequently, is a measure of the water-vapor content of the gas.) Line hydrogen entering the furnaces has a dewpoint of about -10 C. Inside the furnace, the dewpoint may rise as high as 80 C due to oxygen impurities in the hydrogen, leaks or diffusion of air into the furnace, water released by the furnace lining, or simply due to the work load itself. In the last case, the reduction of large amounts of oxides to form water vapor can significantly affect the reaction equilibrium and the final cleanliness of the parts.

Furnace-atmosphere dewpoints, with no work load, should be maintained below 0 C. With countercurrent hydrogen flow (normal procedure), work load and hydrogen flow rate should be adjusted so that the dewpoint at the exit of the hot zone of the furnace does not exceed 10 or 15 C.

Routine testing of firing furnaces consists of checking the furnace atmosphere for small air leaks and for excessive water vapor by observing the surface of a piece of low-carbon-steel strip after it has passed through the furnace. Any discoloration indicates poor atmosphere. Because high-vapor-pressure metals and compounds are volatilized in the furnace, the cooling chamber walls must be cleaned semiannually to control parts contamination.

Organic contaminants are not effectively removed by firing parts in neutral or reducing atmospheres. Some low-molecular-weight organic materials such as methanol, acetone, and acetic acid may be significantly reduced due to their high volatility[5]. Oils, soaps, and other high-molecular-weight substances, however, are

modified by the firing process and generally are made more difficult to remove.

To achieve complete removal of organic materials, it is necessary to destroy the molecule. This is most easily done by destructive oxidation. Firing in air or in wet hydrogen at temperatures of 400 to 1000 C are adequate oxidizing treatments. Subsequent firing in drier hydrogen reduces most of the oxides formed.

Large amounts of contaminants often react with the material of the part during oxidizing to cause stains or other surface defects which are not completely removed by hydrogen firing. To prevent or to minimize this, it is necessary that parts be thoroughly degreased and water washed, or otherwise cleaned, prior to oxidation. Air firing is a "finishing" operation, not a bulk cleaning process. It is the most effective method of removing the organic portions of heavy-metal soaps.

Ultrasonic Cleaning. Ultrasonic agitation is a valuable adjunct to some cleaning processes. It is useful because it can supply energy to detach and disperse contaminants from surfaces. This energy may be used to remove contaminants from places not readily accessible or to permit their removal faster or with less powerful cleaning agents. It is most often used: (1) to remove physically bound particular matter (dust, dirt, carbon), and (2) to aid the dissolving of soluble substances[6]. It is not a cure-all for cleaning problems, however. Often, increase in temperature and flow rate or a reduction in work volume can achieve greatly improved results. In processes like etching, where considerable gas evolution takes place, ultrasonics is of no aid except in speeding up the rate of attack.

The utility of ultrasonic energy in cleaning derives from cavitation and particle movement. Pressure fluctuations can be created in liquids by some electrically or magnetically driven materials. These piezoelectric or magnetostrictive transducers operate at frequencies of 10 to 1200 kilocycles/sec. When the pressure fluctuations exceed the vapor pressure of the liquids, cavitation occurs in the liquid. Cavitation, the formation and collapse of bubbles, provides the scrubbing action important to the removal and dispersion of materials. Low frequencies (10 to 40 kc/sec) are best for cleaning.

Optimum use of ultrasonic methods requires control of solvent flow rate and temperature, of work load, and of basket design.

Chemical Cleaning. This classification includes a large number of chemical solutions used to clean specific materials or to achieve special effects. Only the most generally applicable processes will be considered here. For procedures for cleaning, electroplating, or stripping specific metals, see Ref. 7.

Acid etching solutions are used for some nickel cathode sleeves. Certain high-voltage rectifier types have used bright etching to achieve very smooth anode surfaces to prevent arcing. Electropolishing or ball burnishing may also be used to attain smooth surfaces. For bright etching, the nickel parts are immersed in a heated mixture of nitric, acetic, and hydrochloric acids (usually a 1:1 mixture by volume of nitric and acetic acids with 1-1/2 per cent of hydrochloric acid added after heating the mixture to 70 C). Immersion times are less than one minute. The wall thickness and weight of the cathode sleeves are reduced by this etch. Surface metal and, therefore, surface contaminants are removed. The bright surface and thinner wall increase cathode temperature and appear to offer advantages in some tube types.

Weight loss is, however, difficult to control in this etching process. The gas content of bright-etched cathodes is higher than that of unetched cathodes. Since the occulded gas apparently prevents uniform wetting of the etched surface, rinsing is more difficult and the possibility of contamination from residual nickel salts is increased.

Electropolishing is another method for achieving bright, smooth surfaces. Parts are treated anodically in concentrated acids (usually sulfuric acid, or phosphoric acid, or both) at high current densities. The process allows more control than bright etching but requires that electrical contact be made with each part. The shape and location of the cathode must be closely controlled to prevent shadowing. For these reasons, the process is more suitable for treating large parts and nickel cathode base strips than for treating formed cathode sleeves.

Hydrogen peroxide-formic acid mixtures are versatile cleaning agents for a variety of materials[8]. Formulations can be made to clean or etch, or both clean and etch, such materials as nickel, copper, steel, germanium, and molybdenum. On nickel, solutions of about 3 to 5 per cent peroxide clean with negligible etching and produce a rough matte finish which improves cathode coating adherance.

A twenty per cent acetic acid solution containing 0.2 per cent nitric acid has been used as an effective cleaning agent for kinescope cathode sleeves. A study of the rinsibility of the acid from sleeves is reported in Ref. 6.

Mechanical Cleaning. In the ball-burnishing process, steel balls of an appropriate range in size, together with various soaps or detergents, are placed in a steel drum containing the parts to be treated. The drum is rotated on its longitudinal axis to cause a sliding contact between balls and parts. By the use of proper choice of ball sizes and drum speeds, even fragile cathode sleeves can be brought to a bright (or dull) finish without significant distortion. Bright surfaces can be obtained by ball-burnishing without reduction in wall thickness. The possible improvement in brightness, however, is not as great as that obtainable by acid etching. Ball burnishing has two important disadvantages: very fragile parts may be damaged, and there is danger of imbedding minute particles of foreign materials in the surface of parts.

PROCESS AND MATERIAL CONTROLS RELATED TO CLEANING

Concern with achieving clean parts for electron tubes involves a number of areas. These include:

1. Purity of raw materials

2. Kinds of lubricants used in forming parts

3. Storage conditions from raw material to finished mount

4. Handling conditions, primarily those which follow the final cleaning operations

Any of these areas are possible sources of contaminants. Trouble-shooting tube shrinkage caused by contamination generally requires a knowledge of the processing, storage, and handling history of the tube parts.

Raw Materials

Chemical contaminants of the order of parts per million may result in bulb deposits, leakage paths, or emission poisoning. To minimize the occurrence of chemical contaminants, it is advisable to hold frequent consultations and to collaborate closely with vendors in addition to performing the usual routine chemical and physical analyses of raw materials. Restriction on plating-bath additives and modification of drawing and rolling conditions are examples of process controls which have been required of vendors. Proposed surface coatings or other protective finishes must be carefully examined for sources of contaminants. It may also be necessary to specify the lubricants suitable for use by vendors.

Lubricants

Lubricants and lubrication are of great importance in any consideration of parts cleaning. Their importance stems from the fact that a good lubricant is, almost by definition, difficult to remove. The primary purpose of a lubricant is to prevent metal-to-metal contact during a forming operation. As a result, the best lubricants will, in general, have the most affinity for metallic surfaces. Because factors such as the type of metal, fragility of the part, or limitations on surface conditions may severely restrict cleaning methods, this affinity, often accomplished through some degree of chemical reactivity, may be the source of a cleaning problem. For the most part, a compromise is arrived at to achieve reasonably good lubrication and satisfactory removal. To minimize the number of problems in this area, very strict procedures have been established to assure that proposed lubricants are fully examined and approved before use, and that the number of approved lubricants is held to a minimum. Lubricants are examined for ease of removal and for storage stability with various metals at ambient and higher temperatures. Each approved lubricant is described and specified in an individual standardizing notice [9, 10]

Storage and Handling

Atmospheric pollutants, particulate matter, and packing materials and techniques often are insidious sources of contaminants. Atmospheric pollutants can cause tarnishing or other corrosion of metals, deposition of organic films, or other deterioration of cleanliness that may become evident in tube performance. Particulate matter, such as lint and dust, may create similar problems. Packing materials and packaging techniques may introduce chloride-, sulfur-, or oil-bearing substances in addition to lint and dust. Except where practically irreversible changes have occured, however, the major concern is the storage and handling of tube parts and mounts after final cleaning and before exhausting. The two problems are: (1) how to maintain, during the storage period before assembly, the cleanliness achieved by washing and firing the parts, and (2) how to assemble the parts without introducing contaminants.

Relatively little work has been done on the problems of maintaining clean parts. Before the water-break test was adopted, routine detection of minute amounts of contaminants was impractical and of incidental concern. With the introduction of better cleaning methods, this problem has assumed greater importance. Defining "clean" as the absence of hydrophobic organic materials detectable by the water-break test, Feder, Craft, and Koontz[4] described their results in protecting parts from the atmosphere by sealing in containers.

Their results were:

Polyethylene bags	Failed test* after 1 day
Greased ground-glass joint	Failed test* after 1 day
Petrie dish	Failed test* after 3 days
Screw-cap vial	Failed test* after 10 days
Ground-glass joint (no grease)	Failed test* 15-30 days
Sealed glass tube	Passed test* at 60 days (max. test period)

Most of the containers tested by Feder, et. al. are impractical for high-volume production storage; however, the results indicate the importance of minimal air movement, and point the way toward improving existing containers by using tighter-sealing covers, wrappings or both.

Contaminants in the atmosphere surrounding tube components during storage and handling can greatly affect their performance. Such contaminants may be chemicals such as sulfur compounds, water vapor, plating-room fumes, hydrochloric acid fumes; or they may be particulate matter such as lint, dust, and particles from grinding or forming operations. Here again, the storage and processing history of tube components is necessary to aid in detecting and minimizing or eliminating sources of contaminants.

A significant source of contamination is the tube assembler. Perspiration from hands will cause defective tubes. Hand lotions are another source of contamination. They invariably contain oils and salts which are harmful to tube performance. To minimize con-

*RCA tests show nickel strip no longer water-break-free in less than one hour.

tamination from these sources, it is necessary that finger cots be worn by all persons who regularly handle parts after the final cleaning operation.

SUMMARY

Successful tube making depends on obtaining and maintaining clean parts and, therefore, requires concern with practically all phases of parts manufacture and assembly. Both the amount and the kind of contaminants must be controlled.

In its most general sense, clean means freedom from unwanted materials. Its achievement requires first, that contaminants be removed from the part and, second, that they be prevented from redepositing on the part.

There are five classes of contaminents described: physically adherent, chemically adherent, water soluble, solvent soluble, and gases. If present in tubes, they can cause a variety of defects. Their presence on parts may be established and measured in several ways, the most usual are: (1) microscopic examination, (2) residue tests, (3) water-break test, (4) conductivity test, (5) chloride test, and (6) vacuum-extraction test.

The most widely used cleaning procedure for electron-tube parts consists of degreasing with trichlorethylene solvent followed by washing with deionized water and by hydrogen firing. Ultrasonic agitation of washing liquids, and air oxidation of metal parts are increasingly important cleaning aids. Chemical cleaning has been, and still is, an important aspect of cleaning to meet special situations.

To assure clean parts, it has been necessary to establish controls over a number of phases of parts making. Raw-material purity, forming lubricants, storage conditions and environments, and handling conditions have all required some controls.

ADDITIONAL CLEANING PROCESSES

The preceding treatment of the subject of parts cleaning and processing is oriented primarily toward the processes used in handling large-production runs of products of the type manufactured by the Harrison organization. The following material, by L. P. Fox of the Lancaster organization, discusses additional processes that are used in Lancaster.

Alkaline Electrocleaning

Alkaline electrocleaning is used primarily prior to electroplating. There are a wide variety of proprietary products on the market for this purpose, most of which will do an equally satisfactory job. The alkaline cleaner provides a wetting and emulsifying action to remove organic residues. In addition, when dc voltage is applied between the parts and a steel electrode, the water of the solution electrolyzes and liberates oxygen at the anode and hydrogen at the cathode. These gases rise to the surface of the solution agitating it and accelerating the removal of dirt particles from the metal parts. The process usually produces a surface which is chemically clean and ready for electroplating.

Emulsion Cleaning

Emulsion cleaning is used primarily for heavy-duty cleaning and provides physical cleanliness only. It is usually followed by an alkaline cleaner to obtain a chemically clean surface. There are two types of emulsion cleaners designated as stable- and diphase-type emulsions. The stable emulsion cleaners consist of hydrocarbon liquids such as naphtha or kerosene emulsified with water by means of pine oil, potassium oleate, or other emulsifiers soluble in the hydrocarbon liquids. The emulsions are applied hot (120-160F) by dip or spray, and are followed by a spray rinse in water to remove the remains of the cleaner and the soil. In the case of the diphase-type emulsion cleaner, the metal surface is subjected to the simultaneous action of water and solvent, neither phase being completely or permanently emulsified in the other. The solvent dissolves oils and wets the metal while the water dissolves water-soluble materials. This process, too, is followed by a water spray.

Tumbling

Tumbling is a convenient process for removing burrs or crests which are too large for removal with electropolishing. In addition, this process has the advantage of being able to handle large quantities of small parts which cannot be handled efficiently by the electropolishing method. The process involves placing the metal parts into a barrel containing soapy water, which acts as a lubricant, and an abrasive material, usually aluminum oxide. The parts are rotated, or tumbled, together for a period ranging from two hours up to three or four days in order to remove the burrs and smooth the surface.

REFERENCES

1. Code A1
2. Morgan, O. M. and J. G. Lankler, "Evaluation of Metal Cleaning Compounds," Ind. Eng. Chem., Anal. Ed., v. 14, pp. 725-26, 1942
3. Code A2
4. Feder, D. and D. E. Koontz, "Detection, Removal and Control of Organic Contaminants in the Production of Electron Devices," ASTM Symposium on Cleaning Electronic Device Components and Materials, Special Technical Publication No. 246, 1959, pp. 48, 53, 57
5. Code A3
6. Code A4
7. Code A5
8. Koontz, D. E., et al, The Preparation of Ultra-clean Electron-Tube Components by Chemical Etching, (4), pp. 136-145
9. Code A6
10. Code A7

Coatings, Bases, and Basing Cements

L. P. Fox

Lancaster

Many coatings are used in the production of electron tubes. Some are used for structural reasons, and others to improve tube life. This article describes some of the more interesting and important applications. A discussion of bases and basing cements is also included.

INTERNAL COATINGS

Internal coatings can usually be classified as conductive or insulating. Conductive coatings are applied to insulators such as glass or ceramics, and can be of several types.

Graphite is one of the best known internal conductive coatings and is used in most cathode-ray tube types. The coating medium consists primarily of a dispersion of graphite in water; sodium silicate is added as a binder and gelatin, as a dispersing agent. Fig. 1 shows a graphite coating being applied by the brush method, wherein the tube envelope is placed in a rotating chuck and the brush, wetted with coating, is pressed against the glass and gradually withdrawn. In this manner, the entire internal glass surface is made conductive and can act as a collector of secondary electrons. Before a tube can be evacuated, the coating must be baked at 400 to 425 C for one-half hour to remove water of crystallization from the silicate binder and to oxidize the organic dispersing agent, which in this case is gelatin. This process is known as "bakeout." Because the baked coating absorbs carbon dioxide and water vapor readily, the tube must be evacuated within 24 hours of bakeout. If a greater elapsed time is permitted, excessive quantities of gases will be liberated during and after evacuation. Fig. 2 shows the increase in gas content of the graphite coating with time between bakeout and exhaust.

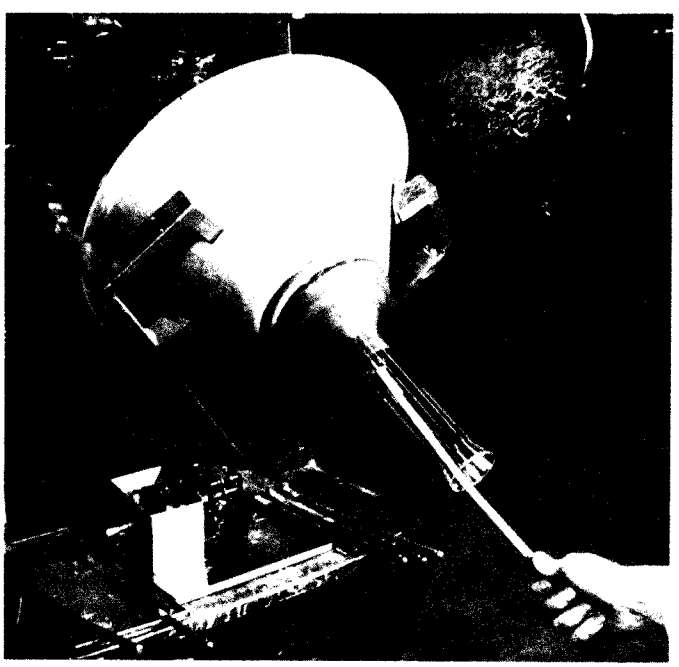


Figure 1. Application of Graphite Internal Conductive Coating

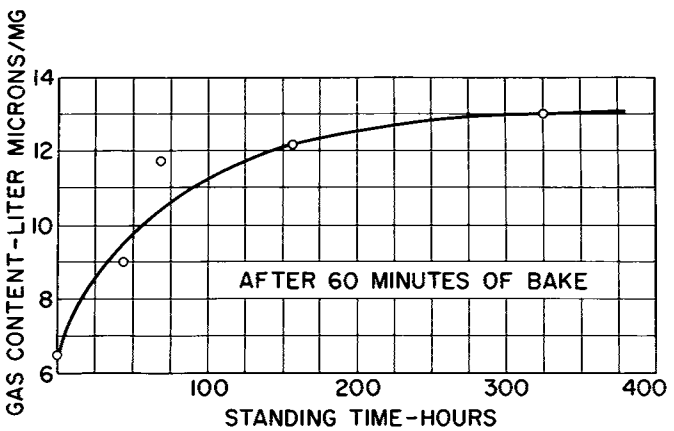


Figure 2. Gas Content of Graphite Coating on Standing

Metal coatings used as internal conductive coatings on glass include evaporated aluminum, silver, and other appropriate metals. Liquid bright platinum, gold, or silver coatings are dispersions of the compounds of these metals in an organic medium. These dispersions are applied to glass by brush and baked at approximately 300 C. They produce a smooth, continuous metallic deposit on the glass surface. Tin oxide is another conductive coating for use on glass. This material is applied by spraying a solution of stannic chloride in methanol on the hot glass at temperatures very near the melting point of the glass. At these temperatures, the stannic chloride is oxidized to tin oxide which provides an adherent transparent conducting film on the glass. Depending on the composition of the glass and the temperature of application, a wide variety of conductivity ranges may be obtained by use of the tin oxide coating.

A coating of iron oxide is often used on the inside surface of cathode-ray tubes to obtain a suitable voltage

gradient. The method of applying the coating is similar to that used for the graphite conductive coating previously described.

Aluminum oxide-silicate-water mixtures are used to form insulating coatings on mica to prevent surface leakages.

All of the internal coatings described use sodium silicate as the binder. Most other binder materials either break down during bakeout, or produce too high a vapor pressure to be used inside tubes. Organic binders can be used only on external coatings.

EXTERNAL COATINGS

External coatings are used primarily on cathode-ray tubes and can be classified as conductive, insulating, or decorative. Organic binders such as vinyls, nitrocellulose, or epoxy resins are normally used. Conductive coatings are used on glass cathode-ray tubes to form a capacitor, with the glass and internal conductive coating as the remaining elements. Graphite is usually combined with one of the binders mentioned above, to impart the conductivity. The insulating coating used on external glass surfaces of picture tubes is a silicone material which reacts with the surface moisture on the glass and then repels further additions of moisture to the glass surface. This material is needed in humid atmospheres for the glass area between the metal shell and the deflecting yoke of metal tubes, and between the ultor connector and the conductive coating of glass tubes. In dry atmospheres, there is no difference in surface leakage between the coated and uncoated glass. However, when the uncoated glass is subjected to high humidity, surface leakage of the order of 250 microamperes can occur at an applied voltage of 27.5 kilovolts.

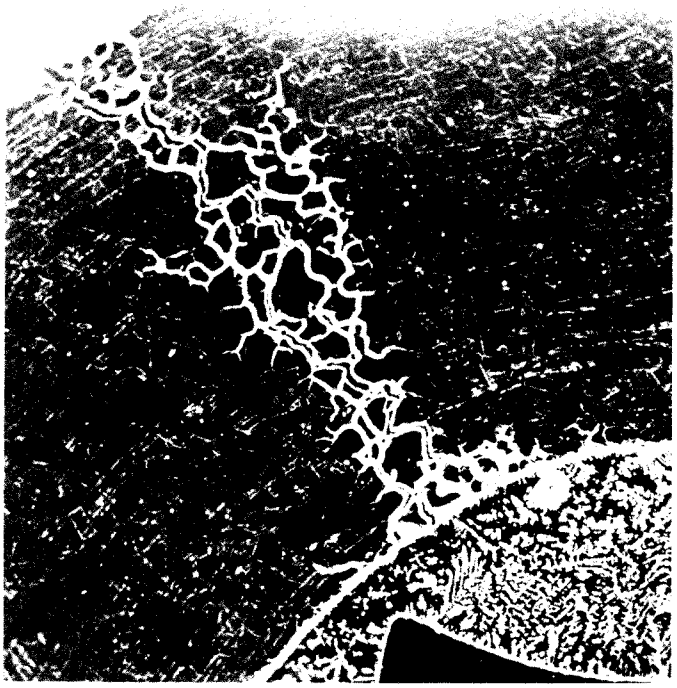
Decorative coatings are often used on metal kinescopes. A black semigloss finish using any of the standard vehicles is often used.

A common requirement for all of the finishes described above is that they must have good adherence to the glass or metal under a variety of atmospheric conditions. In addition, the conductive coatings must have a maximum resistivity of not more than 150 ohms per liner inch when it is measured with a Voltohmyst*, or equivalent, on a dry film 1 mil thick.

METAL COMPONENT COATINGS

Some of the more common electroplated coatings used in the manufacture of electron tubes include nickel, copper, gold, silver, chromium, platinum, bright silver, and bright nickel. These coatings are used for various reasons; e.g.,: copper is used for conductivity; gold, for conductivity and also for its superior secondary-emission characteristics; silver, for conductivity and appearance; chromium, for glass-to-metal seals; platinum, for its secondary-emission characteristics; and bright silver and bright nickel, for appearance. In addition, nickel is used to protect Kovar against corrosion and intergranular silver-solder penetration. Fig. 3 shows the results of the application of silver solder to Kovar with and without the intermediate plated layer of nickel.

(A)



(B)

Figure 3. Effect of Nickel Plating of Kovar on Intergranular Penetration of Silver Solder: (A) nickel-plated Kovar, (B) unplated Kovar

*Registered Trademark

PROTECTION OF SILVER-PLATED TUBES

A method for protecting silver-plated tubes against tarnish during storage and operation has long been sought. Many types of paper wrappers and silver bags have been used in the past, but none has been completely satisfactory because the tubes begin to tarnish from sulfides in the atmosphere as soon as they are removed from the wrappers. A recent commercial chromate surface treatment called "Protectox" has been tested which appears to answer the shortcomings of the previous attempts at tarnish protection. Fig. 4 is a photograph of bright-silver-plated power tubes subjected to normal handling. The first four tubes, reading from left to right, were subjected to normal life testing; the first two were treated and the second two were not. The last two tubes were subjected to an industrial-type atmosphere such as would be found near an acid pickling area. The one on the left was treated; the other was not. This chromate surface treatment is now undergoing further tests to determine whether it can be used on all silver-plated tubes.

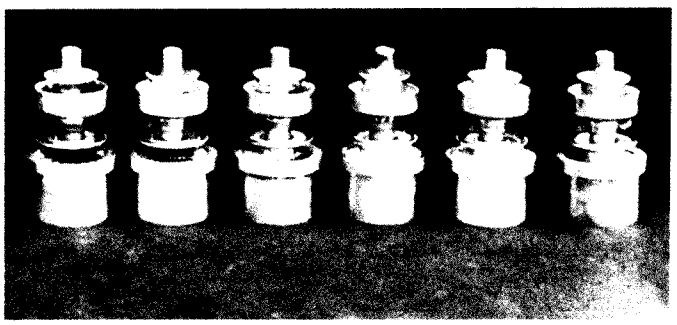


Figure 4. Effect of Chromate Tarnish Protection on Silver-Plated Tubes

BASES

Four types of base materials are currently available. These types are shown in Fig. 5. The base on the left is a general-purpose, wood-flour-filled black phenolic material used on most receiving tubes and picture tubes. A wide variety of grades of this material is used for different applications. For example, picture tubes, which usually operate at high voltages require a base grade having a high dc resistance; receiving tubes are usually not critical to base resistance. The second base from the left is molded melamine, a material of very high dc resistance which, because of its high shrinkage on aging, is not used very much at present. The third base shown is molded of Plaskon, an alkyd-type material used mainly on phototubes or other tubes that require bases that will maintain extremely high dc resistance under adverse moisture conditions. Plaskon has the one disadvantage of being very brittle and easily cracked. The base on the right is made of Micanol, a mica-filled phenolic. It is used for high-frequency applications, particularly on power tubes. Micanol has extremely good dc-resistance characteristics; it is, however, extremely difficult to mold and is more expensive than the other materials described.

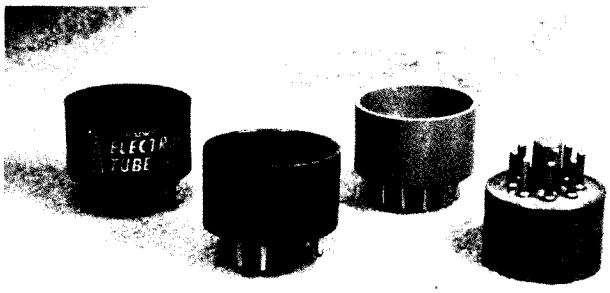


Figure 5. Types of Bases (from left to right: phenolic; molded melamine; Plaskon; Micanol)

BASING CEMENTS

The basic constituents of the adhesive used almost throughout the entire electronics industry are shown in Table I.

Table I

Constituents of Base Cement

Phenolic Resin	Denatured Alcohol
Shellac	Malachite Green
Rosin	Silicone Resin
Marble Flour	

The phenolic resin, shellac, and rosin combine to give the cement its adhesive properties; the marble flour is merely a filler to impart proper flow characteristics and to strengthen the cement; the denatured alcohol is the solvent for the resins; the malachite green is an indicator which shows whether or not the cement has been cured properly; and finally, the silicone resin is added to impart water repellancy to the cured cement.

The final mixture, which is a paste, is extruded by machine into the bases of most high-production tube types. Some bases are handfilled with a spatula by a skilled operator, who can very often apply a uniform quantity of cement to many bases to within several milligrams.

The cement described above is a thermosetting material which means that heat is required for cure and that the reaction is not reversible. In other words, once fully cured, further heat of the same magnitude will not soften the cement. This characteristic is of extreme importance on electron tubes having hot cathodes, because the bases may reach operating temperatures of 80 to 115 C.

REQUIREMENTS FOR GOOD BASE ADHERENCE

Several requirements must be met if good base adherence is to be obtained. The main ones are:

1. The cement must be fresh; maximum storage time for cement is 48 hours.

2. Filled bases should be used 4 to 24 hours after filling unless stored in refrigerated room to minimize alcohol evaporation.
3. The correct amount of cement to fill space between base and glass must be used.
4. Correct curing procedure must be used as follows: (Cement should have light tan color after correct cure.)
   a. Cement must be cured for 1 minute at 150 to 190 C.
   b. The base must be cooled to 90 C before removal from basing head. The bond can be destroyed if the cement is disturbed too early.
5. Bases must not be baked prior to use. Dried bases pick up moisture from the air readily and expand, causing a stress at the cement-base interface.

The base immersion and torsion test is used as a quality check to determine base adherence in a representative sample of production. It is run by soaking based assemblies in deionized water at 50 C for 42 hours. After drying at room temperature for one hour, the base is subjected to a torque, the magnitude of which depends on the base size. Factory limits are as follows:

Diameter of Base-Inches	Factory Torque Limit inch-pounds
1.5 or smaller	40
over 1.5	60
2.0	100

This test does not correlate very well with field-failure results. During field exposure, the tubes are subjected to alternate wet and dry, hot and cold conditions, but never to a soaking such as they receive in this test.

At the time of this writing, work is being done to try to develop an accelerated laboratory base-adherence test which would correlate more closely with field results. Fig. 6 is a photograph of the types of failures

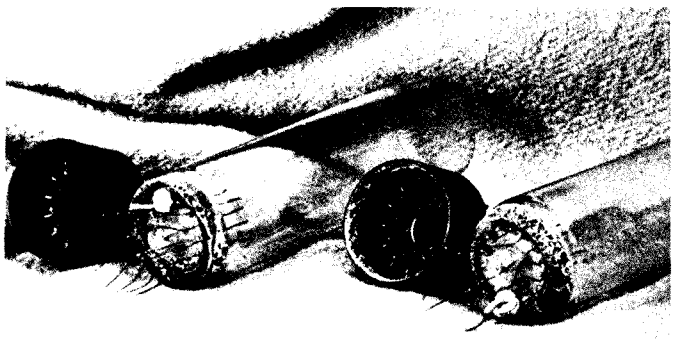


Figure 6. Types of Base Failure

obtained in the field. On the right is an example of a cement-to-cement failure where the adherence of the cement to the glass and to the plastic base is good. This type of failure is usually an indication of a weak structure resulting from improper cement mixing or improper curing of the cement. The failure shown on the left is more typical of the type of field failure experienced, that is, a loosening of the base-to-cement bond. Before the silicone was added to the cement, the predominant type of field failure was cement to glass. The present type of failure indicates that the cement no longer absorbs as much moisture from the air. Because of the addition of silicone to the cement, the base has now become the weak member. In fact, it has been shown that baking, or drying, of the base prior to basing can cause field failures of the cement-to-base type.

Another possible method of improving the life of cement adherence in the field is to develop a new and better cement. There are many new adhesives today in the epoxy, rubber, polyester, and polyurethane resin class. However, up to now, none have proven completely satisfactory from a manufacturing as well as a field-adherence standpoint.

High-Vacuum Technology and Equipment

W. G. Henderson and A. L. Lucarelli

Lancaster

The subject of high vacuum is so broad that full justice to it cannot be done in a single article. It is the hope of the writers, however, that the material which follows will enlighten the layman and serve as a refresher for those who have infrequent contact with the subject.

The use of vacuum as a processing tool has grown tremendously since World War II, and its impact has been felt in all phases of engineering. Although no single industry covers the entire vacuum field, the electron-tube manufacturer is one of the largest users of high-vacuum equipment. At RCA Lancaster there is a wide application of exhaust facilities. Without belittling other operations, the evacuation process can be considered the heart of tube manufacturing. This paper is primarily concerned with vacuum as employed at the Lancaster plant.

WHAT IS A VACUUM?

As defined by the American Vacuum Society (an organization which holds a yearly symposium for the interchange of information among the users of vacuum processes) a vacuum is a gaseous space where pressures are below atmospheric. In the field of high vacuum the degree of vacuum is expressed in pressure units, eg., 10^{-3} millimeters of mercury (mm Hg) instead of vacuum units, 29 inches of mercury (in. Hg). The latter terms are more familiar to engineers in other fields because of their interest in pressure differences used to permit transfer of liquids, braking, and boiler plant applications.

PUMP RANGE

In the electron-tube industry, the vacuum used for the evacuation processes is as low as achieved by any industrial manufacturer employing vacuum. Fig. 1 will give one an appreciation for the range of vacuum obtained. It is interesting to note that the atmospheric pressure at a height of 95 miles (502,000 feet) is less than 10^{-3} millimeter of mercury which, although slightly below pressures achieved by a mechanical vacuum pump, is appreciably greater than those attained by the fractionating pumps. The position of the class of pumps on the chart indicates the approximate degree of vacuum of which each is capable and the different processes for which each is used.

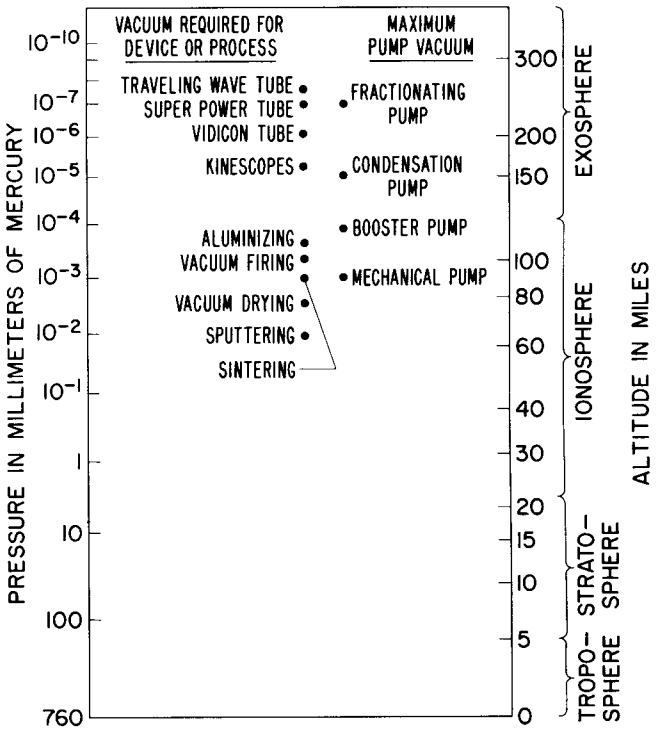


Figure 1. Schematic Cross-Section of the Atmosphere and Relationship to Evacuation Processes

VACUUM CLASSIFICATION

Table I presents a tentative[2] classification chart for ranges of pressures expressed in mm Hg, microns, and microamperes of current as read with an RCA 1949 ionization gauge tube which has a sensitivity of 100 to 1. It makes clear the relationship between microns and millimeters of mercury and introduces the term "torr." Torr (short for Torricelli, an early vacuum experimenter) is a term which, although not used in our organization, may be encountered. One micron is 10^{-3} mm Hg, for using an ionization gauge control equipped with an RCA 1949 tube; a micron is equivalent to a reading of 100 microamperes.

TYPES OF FLOW AND MEAN FREE PATH

In the pump-down of a system, the flow of gases

Table I

Vacuum Classification Chart

Range	Pressure mm Hg	Pressure Microns	RCA 1949 Gauge Current Microamperes
Medium Vacuum	1.0 to 10^{-3}	1000 to 1	--- to 100
High Vacuum	10^{-3} to 10^{-6}	1 to 0.001	100 to 0.1
Very High Vacuum	10^{-6} to 10^{-9}	0.001 to 0.000001	0.1 to 0.0001
Ultra-high Vacuum	10^{-9} and lower	0.000001 and lower	0.0001 and lower

Conversion of Units

Microns	mm Hg	Torrs	Microamperes*
1000	1.0	1.0	100,000
1.0	0.001 or (1 x 10^{-3})	0.001 or (1 x 10^{-3})	100
0.01	0.00001 or (1 x 10^{-5})	0.00001 or (1 x 10^{-5})	1

*For gauges which read 100 microamperes per one micron of pressure.

moves through three flow stages: viscous, transition (or slip), and molecular. In the viscous flow range, which occurs during initial pump-down from atmospheric pressures, the gas laws (Boyle, Charles, and Poiseuille) hold true. The slip flow, which occurs at the lower order of pressures produced by a mechanical vacuum pump, is a transition region running into the molecular flow range. In the molecular range, which starts at approximately one micron, the behavior of gases is different from that in the higher pressure regions. From Avogadro's hypothesis (for a given pressure and temperature gases have the same number of molecules per unit volume), it is evident that a smaller number of molecules will exist in a chamber at reduced pressures. The kinetic theory states that these molecules are in random motion. It follows that as the pressure drops off and the number of molecules decreases, each gas particle must travel a longer distance before colliding with another. The average distance a molecule must travel before colliding with another is known as the "mean free path" and is of extreme importance in the successful operation of evacuation process. The mean free path for air at room temperature expressed in practical values for different pressures is shown in Table II. Note that the mean free path varies directly with the pressure.

SYSTEM COMPONENTS

The system used to attain low pressures is illustrated schematically in Fig. 2. Depending on the process to be carried out, there may be duplication or deletion of units. For example, in vacuum drying or in some types of vacuum firing a single mechanical pump may suffice, whereas, in a high-speed, high-vacuum application several traps, mechanical pumps, high-vacuum pumps, three types of pressure reading instruments, and a multiplicity of valves may be required. The attainment of the lower order of pressures (below 10^{-5} mm Hg) requires that gases from the tube must be made to flow into a high-vacuum pumping system capable of attaining the desired ultimate pressure. Not illustrated, but important to the successful operation of the tube exhaust, are the means for accelerating the liberation of gases from the envelope and elements of the tubes. The equipment employed for such purposes includes: (1) bake-out oven for heating the glass or metal envelope to drive off adsorbed and absorbed gases from the tube walls; (2) radio-frequency bombarders to outgas internal metal elements; and (3) circuits for activating the cathode of the tube by filament, or heater power, to establish the cathode as a satisfactory emitter.

Table II

Mean Free Path
(Air at 25 C)

Pressure—mm Hg	Mean Free Path—Cm
760	0.0000067
10^{-2}	0.5
10^{-3}	5.0
10^{-5}	500.0

GETTERING

Generally, after evacuation and "tip-off of the tube, a gettering process is performed to remove additional gas molecules present. The process involves exposing the gases to absorption surfaces on which the gases impinge and are held by sorption or by chemical forces.

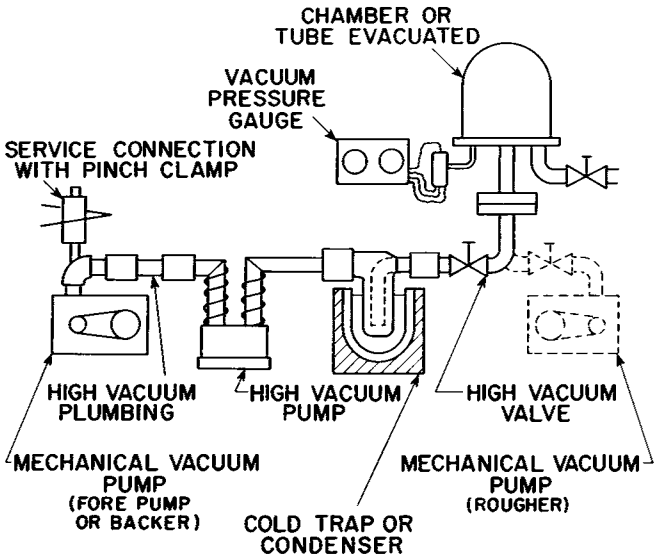


Figure 2. Vacuum System Schematic Diagram

In several large power tubes, for example, an overheated anode, in cooling down, will getter residual gases; in other tube types a barium compound is flashed to chemically react with the residual gases to form compounds which settle on tube surfaces. This gettering action reduces the ultimate pressure in the tube by a factor of 10 to 100. The improved degree of exhaust minimizes the destructive ion-bombardment effect on the barium cathode surface.

REQUIREMENTS FOR A GOOD VACUUM

In addition to using adequate components, a system must be free of contaminating materials if low pressures (10^{-6} mm Hg or lower) are to be successfully attained.

The partial pressure of the gas and vapor evolved from materials exposed in a vacuum will usually depend on the area of material exposed, the nature and previous treatment of the material, and the system pumping speed. Because practical considerations often limit the size of exhaust piping, more can sometimes be accomplished by proper selection of materials to be exposed in the system than by increasing the pumping speed. Therefore, in the design of a system, good vacuum practice dictates the use of materials having low out-gassing rates, selection of suitable components, correct plumbing (flanges, pipe, manifolds, rubber or metal couplings, and castings), and choice of proper welding techniques. As an example, seamless tubing must be used instead of commercial welded pipe to avoid possible leaks occasionally encountered with use of the latter. Also, improper brazing always proves troublesome due to high out-gassing from excessive flux usually employed. More mention will be made later regarding contamination and sources of trouble.

MECHANICAL PUMPS

For an understanding of the operation and function of components shown in the schematic pumping system of Fig. 2, it is well to start with the mechanical vacuum pumps. The primary function of a mechanical pump is to reduce the pressure from atmospheric pressure to pressures below 100 microns or, more often, below 10 microns. These pressures are suitable for some processes (see Fig. 1). More important, however, they are required for successful operation of the high-vacuum pump. Fig. 3 depicts one pump from each of three leading manufacturers. The Kinney KC-15 (from Kinney Manufacturing, Boston, Mass.) and the Welch 1402 (from Welch Scientific Co., Chicago, Ill.) are the most popular pumps. All mechanical pumps are rotary, positive-displacement pumps with single or compound stages. Their capacity, or pump speed, is expressed in free air displacement (cubic feet per minute, cfm, or liters per second, l/sec). Compound pumps are preferred over single-stage types because of their ability to produce lower pressures which, in turn, contribute to attainment of lower ultimate pressures which, in turn, contribute to attainment of lower ultimate pressures in the system. Most mechanical compound pumps are capable of attaining pressures from 1 micron to 0.1 micron when received from the manufacturer.

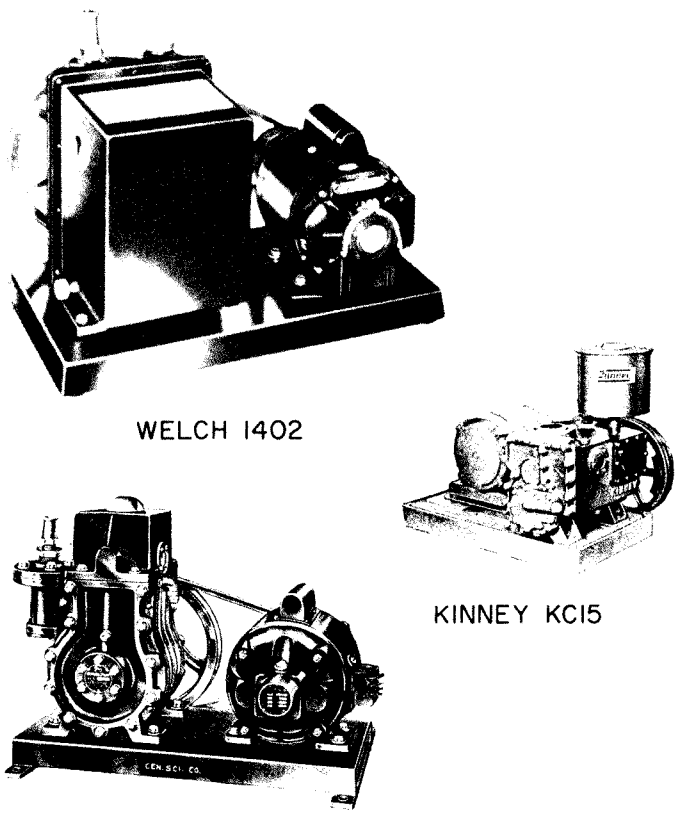


Figure 3. Typical Mechanical Vacuum Pumps

Fig. 4 shows a cross section of the Kinney pump. An eccentric rotor wipes the concentric stator to trap a pocket of gas which is ejected to atmosphere through a poppet valve. The valve, a rectangular block with spring-loaded studs, seats over outlet holes drilled in the pump body. When gas is ejected from the pump, the valve unseats. The valve flutter caused by the es-

cape of trapped gas molecules produces a pounding noice characteristic of all mechanical pumps. When the pump is turned on, the large slugs of gas unseat the valve a maximum amount causing a loud pounding which diminishes to a steadier lower noise level as the pressure is reduced. This change in noise level can be used as an indication of a system devoid of large leaks. An initial, persistent, loud pounding should alert an operator to look for a leak, such as an open leak-valve, open head port or open bell-jar port. The mechanical pump should never be allowed to operate for more than 15 minutes under free air-flow conditions because the pump oil will heat up and the hot oil will then become more readily contaminated. During the operation of a pump, successive gas removals eventually cause contamination of the pump oil necessitating a pump service (cleaning and recharging the unit with new oil). It follows that pumps with small oil charges will require more frequent servicing. Hence, in systems where heavy gas loads are encountered it is well to employ pumps with large oil reservoirs.

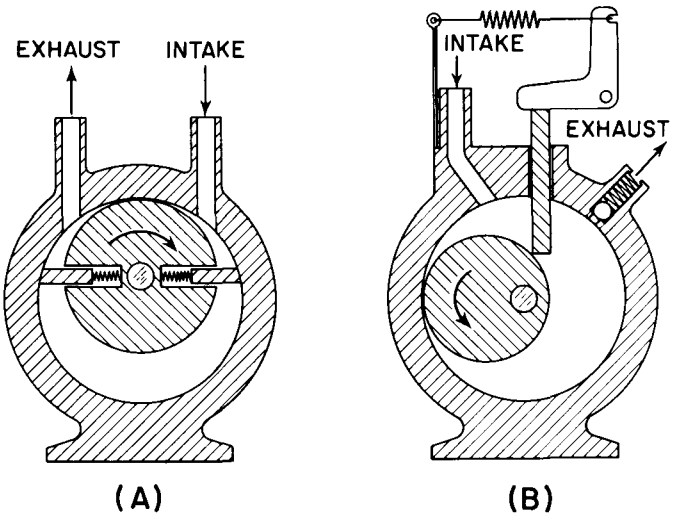


Figure 5. Cross-Sections of the Welch Pump (A) and the Cenco Pump (B)

Fig. 5 shows cross-sections of the Welch and Cenco pumps. The Welch pump houses a concentric rotor with spring-loaded vanes which are in continuous contact with the eccentric stator. The Cenco type pump has an eccentric rotor with a spring-loaded nonrotating vane in contact with the rotor.

MECHANICAL-PUMP SPEEDS

The speeds for a variety of rotary oil-sealed pumps, are shown in Fig. 6. The curves show the pump characteristic (speed vs. pressure) from 500 mm Hg to the ultimate (blank off) pressure. These curves, given by the pump manufacturer, are obtained from readings on a McLeod gauge under ideal operating conditions. The curves are for compound pumps except for those of the KD (Kinney Dual) and KS (Kinney Single-Stage) series. (Note: In the Kinney pump number, the figure given is the approximate rated free-air capacity in cubic feet per minute). The advantage of compound pumps over single-stage pumps can be appreciated by observing that the curve for the single-stage pump (KS) drops off at a pressure of about 10^{-2} mm Hg, whereas, compound pumps such as the KC-15 continue on to below 10^{-3} mm Hg. Although the selection of a mechanical pump may be influenced by the equipment arrangement, more often the choice is based on the volume of the chamber to be evacuated, the rate of evacuation desired, and the process involved (the sequence of operations).

EXHAUST TIME — MECHANICAL PUMP

Over a pressure range where the pump speed is nearly constant, the time of evacuation can be determined by

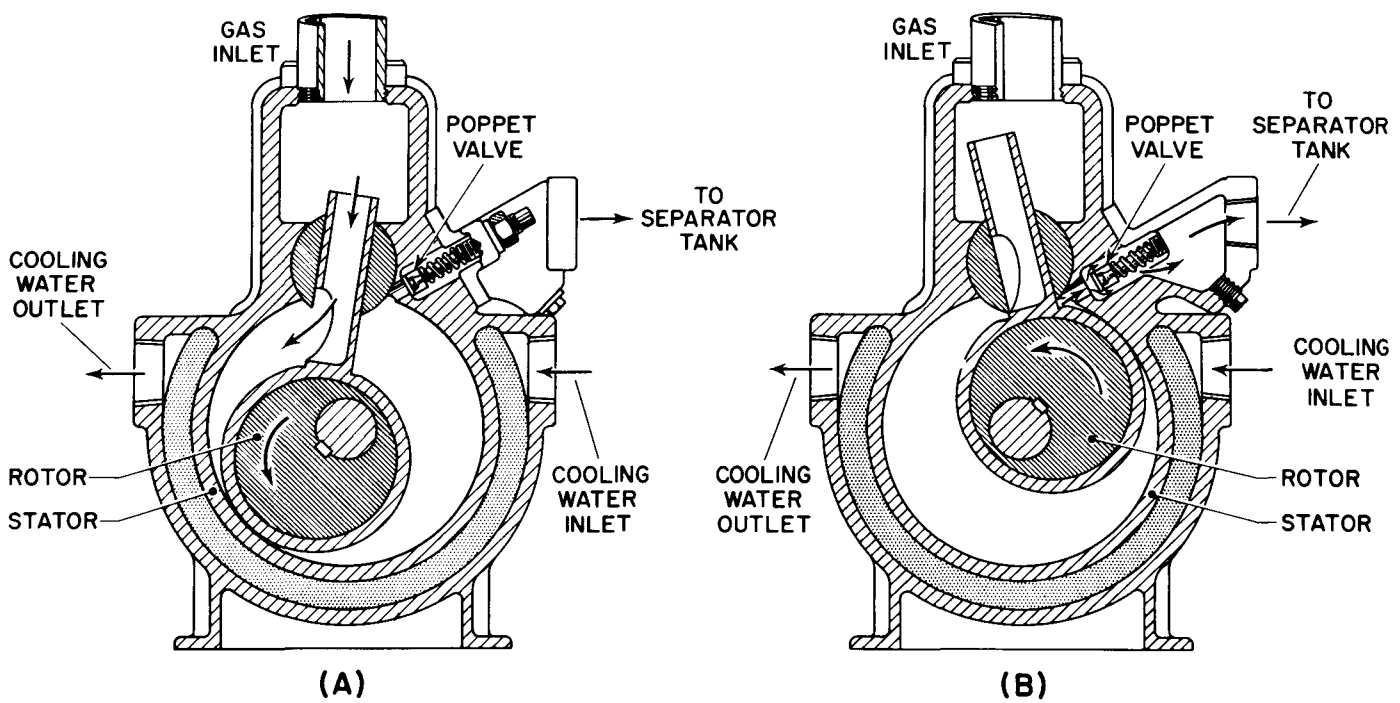


Figure 4. Cross-Section of the Kinney Pump: (A) intake; (B) discharge

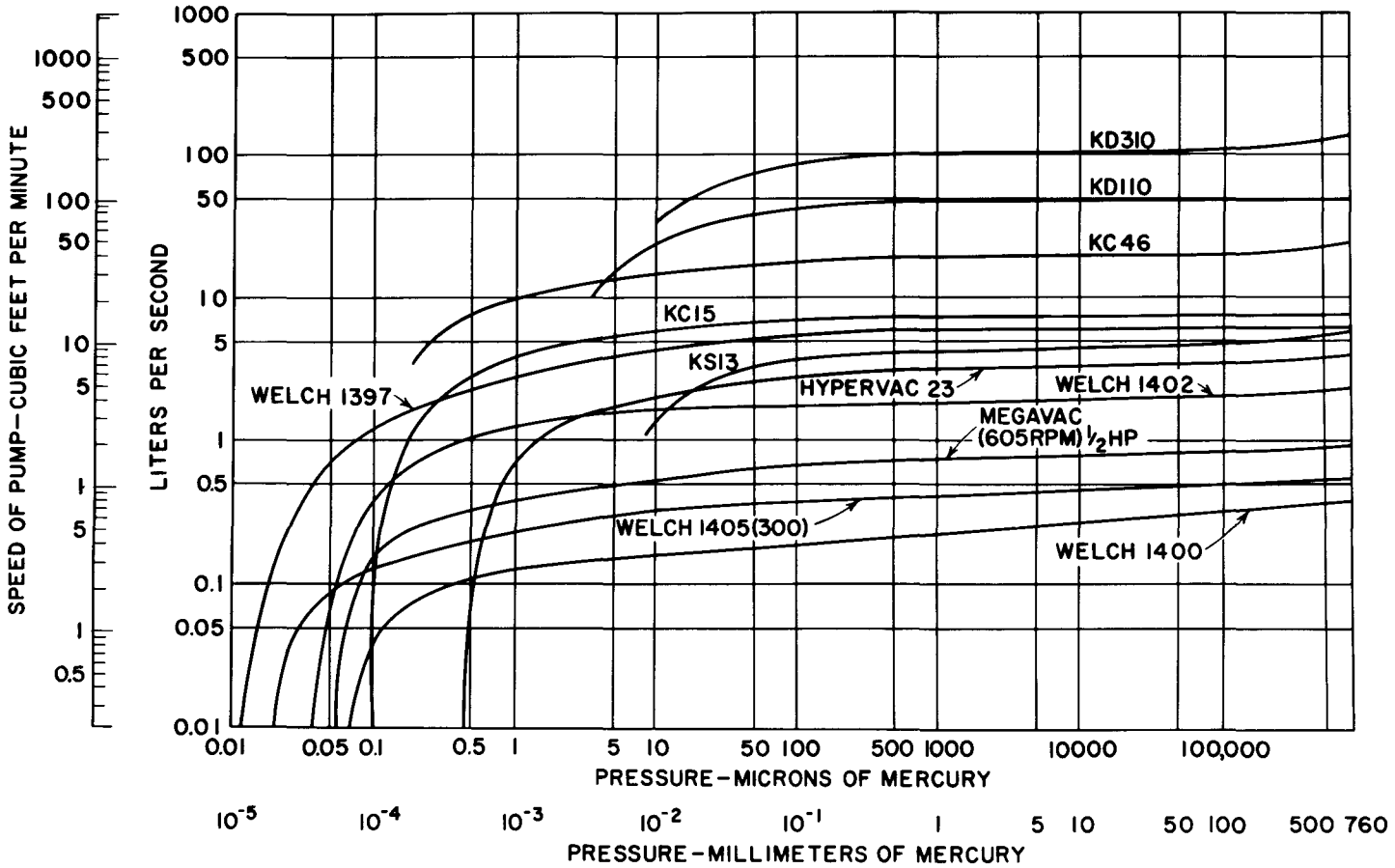


Figure 6. Pumping Speeds of Rotary Oil-Sealed Pumps

the following equation:

Time of evacuation from P_1 to P_2,

$$t_2 - t_1 = \frac{2.3V}{S_a} \log \frac{P_1 - P_2}{P_2 - P_u}$$

where S_a = Average pump speed (assumed constant)
V = volume of system
P_1 = pressure at time t_1
P_2 = pressure at time t_2
P_u = ultimate pressure (base or lowest pressure attained)

Note: All above units are to be consistent.

This equation is not reliable for low pressures where system out-gassing (evolution of gas) occurs. It is, however, of value in determining time required to rough out a system to 100 to 500 microns. The formula can further be simplified for an approximation of the time T required to rough-out from atmospheric pressure to a pressure of 500 microns (0.5 mm Hg).

$$T = \frac{7V}{S_a}$$

Since the average speed of pumps through their pump-down cycle is generally about 3/4 of the rated speed, the approximate formula to rough-out to a pressure of 500 microns can be written as

$$T = \frac{10V,}{S_r}$$

where S_r is the rated pump speed (free-air capacity) as given by the manufacturer. For example, a KC15 (rated at 15 cfm) connected to a system of 3 cubic-feet volume would "rough" it out in approximately 2 minutes.

GAS BALLAST PUMP

In processes where considerable amounts of water vapor must be handled, a gas ballast pump is recommended. This type has proven its superiority over the conventional type in handling water vapor. Condensation of the water vapor in a conventional pump will readily contaminate the pump oil to create high-vapor-pressure molecules. The presence of these contaminating molecules will limit the ultimate vacuum obtainable, and may often necessitate complete shut down for reservicing of the pump. Electrical and steam separators, or centrifuges have been employed to combat contamination of the oil. The gas ballast pump, however, is by far the best solution to the problem. The Gaede gas ballast pump, manufactured by Leybold of Germany, was introduced into this country by National Research Corporation. Fig. 7 schematically illustrates the gas ballast feature. By way of the air inlet, air is admitted into the pump throughout the compression cycle. The mixture of air and water vapor is

compressed only to a ratio of 6:1 or 10:1 instead of the 700:1 of the conventional pump. As a result, the vapor does not condense at all, but is discharged with the admitted air through the outlet valve before condensation. A gas ballast pump has been known to evacuate a bell jar system containing a beaker of water without any ill effects to the pump. The ultimate vacuum obtainable with a gas ballast pump is not as low as that obtainable with the conventional mechanical pump. Blank-off pressures of about 10 to 50 microns are common with gas ballast pumps. However, after the initial load of water vapor has been removed, the air inlet valve can be closed and the pump will then continue pumping down to the ultimate vacuum of a conventional pump.

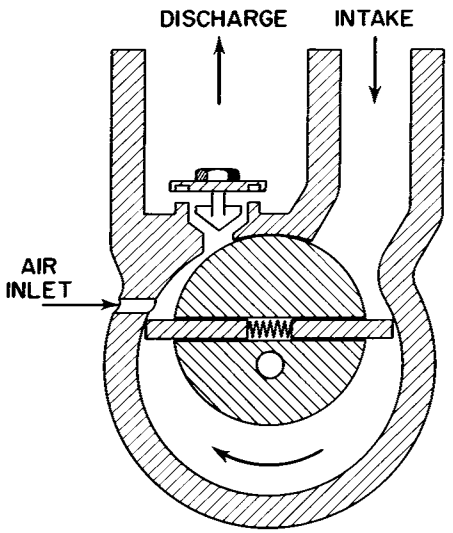


Figure 7. Cross-Section of Gas Ballast Pump

MECHANICAL BOOSTER PUMP

The newest of mechanical pumps currently on the market (at least in this country) is the mechanical booster pump. This unit appears to offer possibilities for use in vacuum firing and other processes performed at pressures of 10^{-4} mm Hg. As shown in Fig. 8, the pump comprises two positive displacement pumps, each with its own motor drive; one is a conventional rotary piston pump and the other is a lobe type pump. In operation, the rotary pump removes the initial slugs of gas to a pressure of about 15 mm Hg, at which point the lobe pump "cuts in" to continue rapid pump-down of the system. During the lobe pump operation, the rotary pump acts as a backer. In the event of a sudden pressure rise the rotary pump automatically takes over as the operating fore-pump. Although the lobe pump may be "cut out," it is possible to arrange for its continued operation during the entire pumping cycle. To do this, however, a torque-limited motor is required, and its control equipment is complex and expensive. Normally, the lobe pump operates at three to four times the rpm of the rotary pump and requires approximately twice the horsepower. The lobe pump increases the pumping capacity by a factor of 8 to 10. As an example, one pump manufacturer lists the following ratings for a mechanical booster pump.

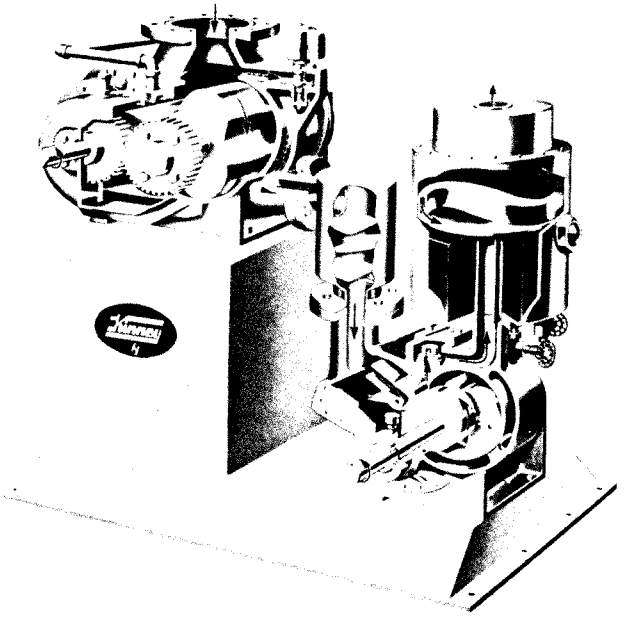


Figure 8. Mechanical Booster Pump

	Rotary Type	Lobe Type
Speed (rpm)	520	1800
Horsepower	5	10
Free Air Capacity (cfm)	131	1230

Fig. 9 shows the performance curve for the above-rated pump. Because both units are positive-displacement pumps, they do not stall and have good volumetric efficiency over a wide pressure range. The lobe pump, in addition to being able to attain pressures of 10^{-4} mm Hg, has no backstreaming of volatile vapors to contaminate the system being evacuated. These features are possible because the lobe pump is operated dry.

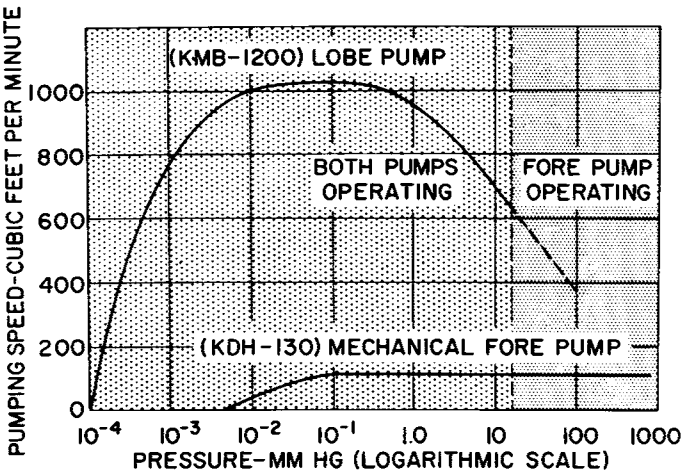


Figure 9. Typical Pumping Speed Curve of a Mechanical Booster Vacuum Pump

OIL-DIFFUSION PUMP

Where considerable quantities of gas must be removed from a chamber at pressures less than 10^{-3} mm Hg, the mechanical pump is inadequate. The diffusion pump is the cheapest unit that can achieve high pumping speeds down to pressures as low as 10^{-7} mm Hg. This pump can be either oil- or mercury-operated. An oil-diffusion pump usually consists of at least two vapor jets arranged so that the vapor stream can propel gas molecules toward its outlet. Fig. 10 shows the operation of an oil pump. Oil in the boiler is evaporated by a heater. The vapor particles, represented by arrows, rise and leave at high speed through the nozzles. Gas molecules, represented by dots, which wander into the vapor path are molecularly bombarded and propelled toward the outlet of the pump for removal by a mechanical pump. Cooling coils in the pump jacket condense the oil vapors, which return to the boiler section for recycling. As the gas molecules are trapped and pushed toward the exit of the pump, a progressively increasing compression of gases occurs. From nozzle to nozzle, the compression may be of the order of 5 or 10 to 1. Regardless of how effective the third (lowest) compression stage may be it cannot eject gas molecules against atmospheric pressure. Therefore, at the exit of the pump, a low fore-pressure must be maintained. The lower the operating fore-pressure, the better the ultimate vacuum of the system. The mechanical pump employed to attain this fore-pressure is known as the fore-pump, or backer.

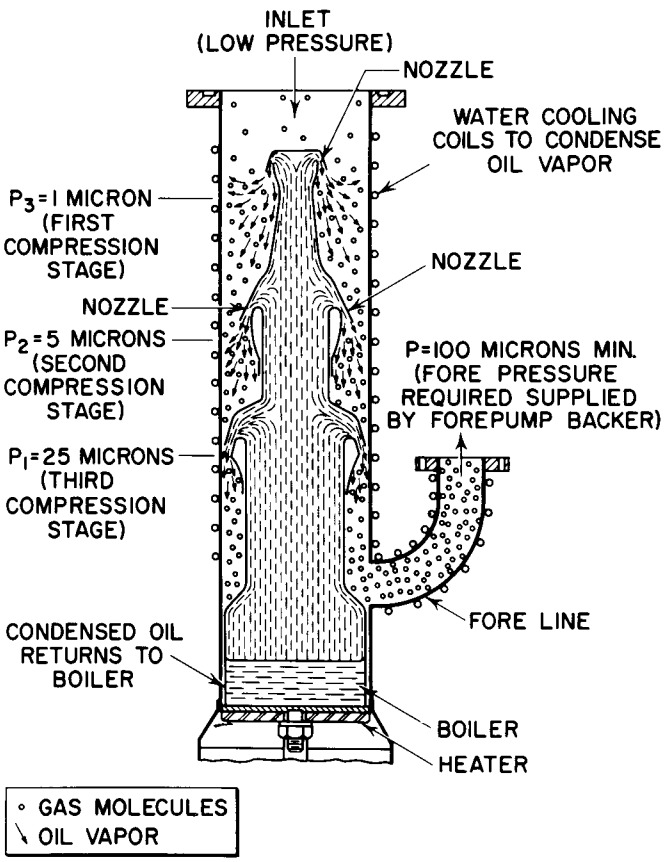


Figure 10. MCF Type Oil Pump

FRACTIONATING PUMP

High-vacuum oil pumps are either fractionating or non-fractionating (condensing). Lower ultimate pressures may be achieved with the fractionating type. The fractionating, or purification, principle is illustrated in the cross-sectional view of the VMF type pump, Fig. 11. This pump is a two-stage unit with a partitioned boiler chamber. The left half of the chamber contains baffles arranged in a staggered pattern. Vaporized oil, when condensed, will return and enter the boiler chamber at the left side. In seeking its own level, the oil will flow through the baffle arrangement and follow a long tortuous path. The heat applied "boils off" the dirty, higher-vapor-pressure oil particles. The cleaner oil fractions find their way to the right half, or high-vacuum portion, of the pump where they are vaporized to produce the higher vacuums in the desired region.

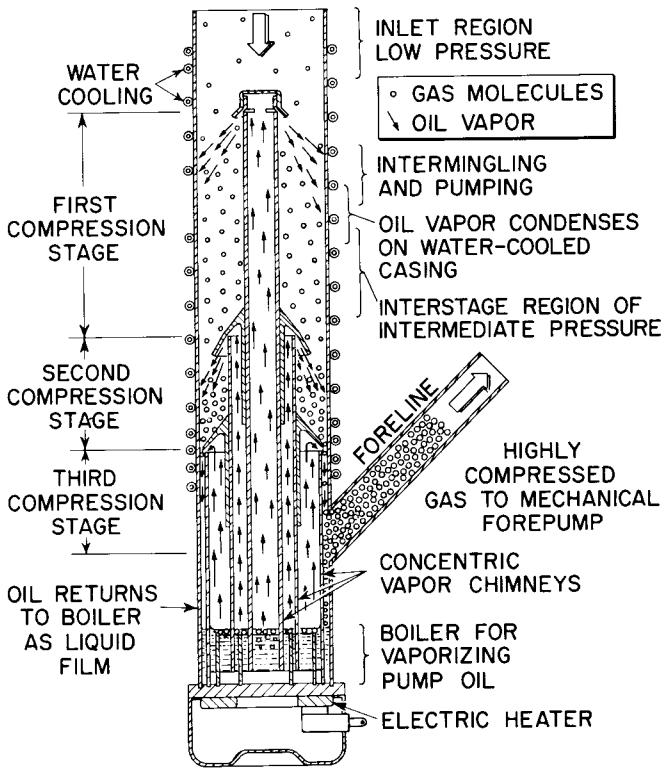


Figure 11. VMF Type Oil Pump

BACKSTREAMING

The most objectionable feature of the diffusion pump is its backstreaming characteristic. Backstreaming is the flow of pump vapors toward the chamber being evacuated. The effects of backstreaming are most noticable immediately above the pump nozzles. For this reason it is best not to attach a pump directly to the vessel to be exhausted. Additional information on reducing backstreaming follows under "Baffles and Traps." A reference article on this subject is available.[3]

The conditions which may exaggerate backstreaming action are:

1. Poor jet design (oversize throat)

2. Excessive pump heating
3. Choked "backer"
4. State of equilibrium

An oversize pump throat (the area between the topmost nozzle and pump casing) will influence the effectiveness of the vapor stream. When this area is excessive, it permits larger quantities of gas molecules to penetrate into the vapor stream region, and, due to kinetic impingement on the gas molecules, a reverse flow can more easily occur. A good jet design, which produces a well-directed and complete vapor stream, helps to overcome backstreaming.

The vapor, as it leaves the jet, has an expanding flow pattern. Excessive heat tends to create eddies at the lower lip of the umbrella cap of the jet; the eddies increase the migration of oil vapor toward the chamber.

A choked "backer," that is, a mechanical pump operating at too high a pressure, or one incapable of removing heavy gas loads, will permit higher pressures in regions below the oil pump jets, and thereby tend to produce backstreaming.

After prolonged pumping, the pressure in a vessel will equal the ultimate pressure of the oil-diffusion pump. At this stage, an interchange of molecules occurs. For every gas molecule that leaves the chamber, an oil molecule will rise from the pump region to maintain the state of equilibrium.

PURPOSE OF CONSTRUCTION

Another, more popular, type of fractionating pump is illustrated in Fig. 12. The chimneys, or stacks, are arranged concentrically which, in addition to simplifying construction, permits the complete jet assembly to be removed for cleaning. Fractionation in this pump is achieved by having the condensed oils circulate around each chimney before entering an inner chimney.

The space between the jets and the jacket wall increases from the lowest to the highest jet. The small spacing at the lowest jet permits the pump to operate against high fore-pressures at low speeds; the large spacing at the highest jet permits high speed against low fore-pressure. In this manner, maximum volumetric efficiency is achieved. The speed of a diffusion pump is the volume of gas per unit of time which passes the throat at a particular pressure. The theoretical maximum flow (over a constant speed range in liters per second) is 11.7 times the clearance area in square centimeters. However, due to thermodynamic inefficiencies the actual flow for modern commercial pumps is 25 to 40 per cent of the theoretical rate.

PUMP COOLING

Except for silicone oils, such as Dow-Corning DC 702 and DC 703, all known oils employed as pumping media will become contaminated when operated with leaks in the system. Excessive air admitted to the pump will readily mix with the oil and result in maximum rate of contamination. If the condition prevails for an extended period, the oil may be heated to its

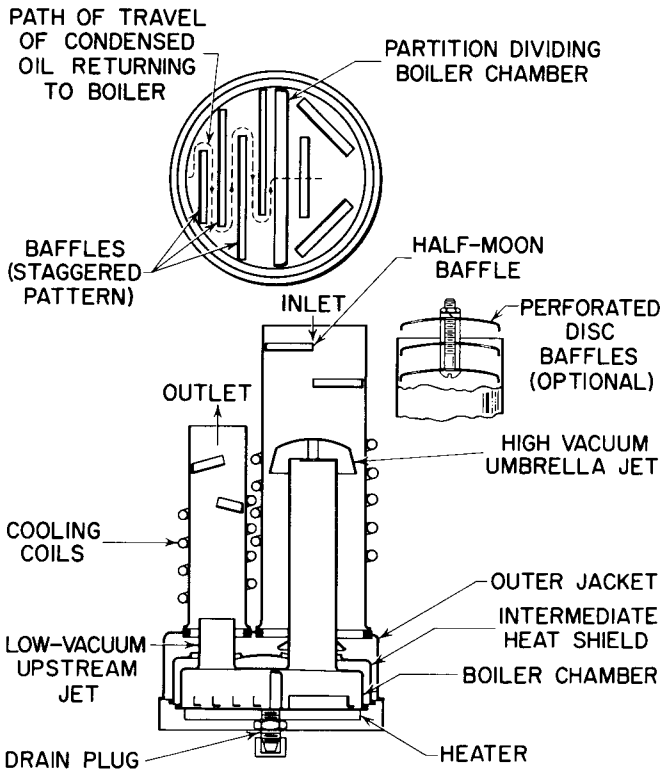


Figure 12. Operation of a Typical High-Vacuum Vapor Pump

cracking temperature. Should cracking occur, harmful deposits will settle not only in the pump but in other components of the system. If cracking temperatures are not reached, the heated oil in the presence of air can become carbonized. In either case, costly maintenance and machine down-time will result. To minimize contamination or breakdown of the oil, the pump heater must be turned off and the pump boiler cooled before the system is opened to the atmosphere. To permit the boiler to be cooled rapidly, many pumps have provision for air or water cooling. The minimum period for cooling the boilers of oil-diffusion pumps with capacities up to approximately 300 liters per second is:

No induced cooling (natural convection) - 15 minutes
Air cooling (high-pressure air or blower) - - - - - - - - - - - - - - - - 4 to 6 minutes
Water cooling (around boiler chamber) - - - - - - - - - - - - - - - - - - 2 minutes

When uninterrupted production must be maintained, pump service should be performed on a regular schedule. Preventive maintenance should be scheduled at intervals of four to eight weeks; in isolated cases the periods can be longer, possibly as much as twelve weeks of 24-hour-per-day operation.

PUMP CURVE FEATURES

The characteristic curve of a pump is always provided by the pump manufacturer. Fig. 13 shows the speed-pressure relationship, under conventional oper-

ating conditions, for three types of diffusion pumps — the booster pump, the condensation pump, and the fractionating pump. The booster pump can operate at the highest pressures of the three, but its ultimate vacuum is the poorest.

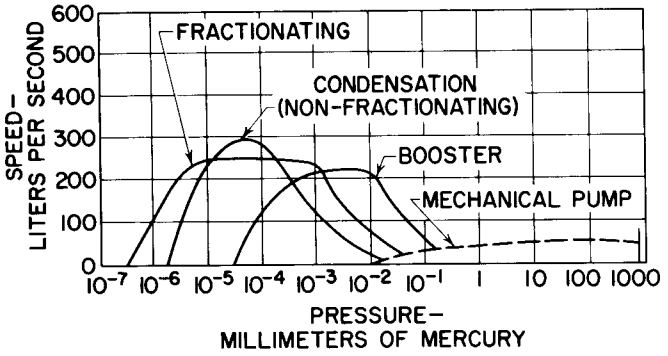


Figure 13. Speed Variations of Oil Diffusion Pumps

The fractionating pump, in addition to achieving the best ultimate vacuum, has a flat peak indicating a high speed over a large pressure range. The high peak of the condensation pump is an arbitrary level and does not imply a top speed greater than that of either of the two other types of pumps. The maximum speed of a pump is dependent on its size, on boiler pressure, and on internal design.

Booster pumps operate against fore-pressures in the millimeter range. The principle advantages of the booster are: (1) it provides maximum speed over a range which bridges the slow-speed gap between the limit of the mechanical pump and the start of the high-speed, high-vacuum pump; (2) it can remove large amounts of gas in the pressure range that satisfies processing needs at the lower order of pressures. For example, in the aluminum evaporation process which requires pressures of 2×10^{-4} to 5×10^{-4} mm Hg, a booster pump with an ultimate pressure of 10^{-4} mm Hg, is preferred. A condensation pump or a fractionating pump offers no advantage for this process because, in the desired operating range, either is only approaching peak speed. Systems requiring pressures of 10^{-6} mm Hg or less must employ a fractionating pump.

The heat input to a diffusion pump affects the pump speed characteristic. Fig. 14 shows curve changes in a typical pump obtained by variations in heater current. Although not indicated, reduced current shifts the start of pump operation to the left, and increased current causes a shift to the right.

In recent years, Dr. R. B. Lawrence developed an improved graphical method of depicting a pump characteristic.[4] Fig. 15 shows a typical set of such performance curves. The normal speed-pressure curve is rotated to a 45-degree axis and the ordinate is expressed as throughput (or gas flow) in micron-liters per second over the indicated pressure range. The curve for a recommended mechanical pump is shown together with the tolerable and breaking forepressure for the oil-diffusion pump. At a glance, the important values can

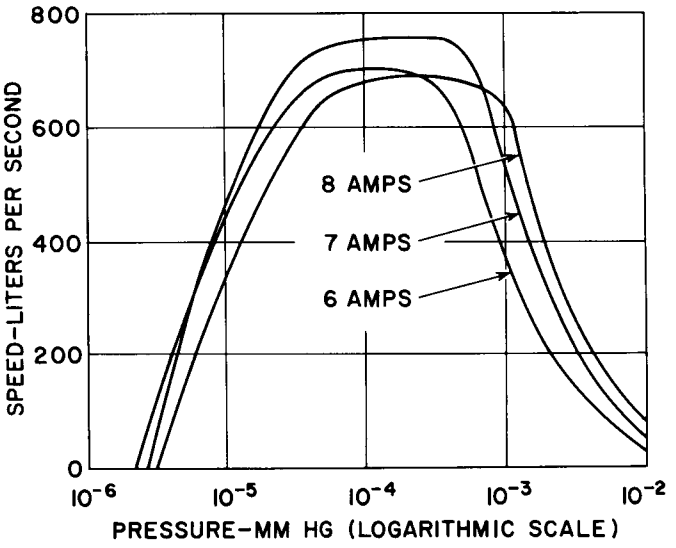


Figure 14. Pump Speed versus Pressure of a Typical Oil Diffusion Pump With Heater Current as a Parameter

be obtained readily. These include blank-off pressure, mid-speed (peak) range, pressure range over which constant speed is obtained, maximum throughput, and fore-pressure characteristics. The horizontal spacing between the diffusion-pump curve and the curve of the typical operating fore-pump represents the compression achieved by the vapor pump. The tolerable fore-pressure curve forms a boundary which indicates that operating fore-pressures to the left are acceptable while fore-pressures to the right become increasingly unsatisfactory. A backer smaller than the pump represented by the typical operating fore-pressure curve can be tolerated. It will cause deterioration in high throughput performance, but will have no effect on the vapor pump performance through the medium-and-low-throughput ranges. On the other hand, a larger fore-pump can improve performance but only within limits. Fig. 16 shows the effect of curve shifts. In general, as the curves move outward from each other, operating characteristics improve.

SYSTEM CURVES

In practice, performance vacuum curves may be used to: (1) check out a system; (2) determine causes of poor system operation; and (3) permit optimum processing procedures to be devised. Fig. 17 shows a representative pressure-time curve for tube exhaust. The hump at a pressure of approximately 100 microns (10^{-1} mm Hg) indicates that the vapor pump is taking effect; the increase in pressure is the result of outgassing from oil in the vapor pump. The degree of outgassing will vary with pumps and is influenced by the boiler heating method, the time in the cycle at which the pump heater is turned on, and the lag between pumping cycles. The second hump may indicate bursts of adsorbed or absorbed gases suddenly released by external causes such as external heating of the tube. The third hump shows the outgassing effects from other tube treatments such as radio-frequency bombardment. In most tube manu-

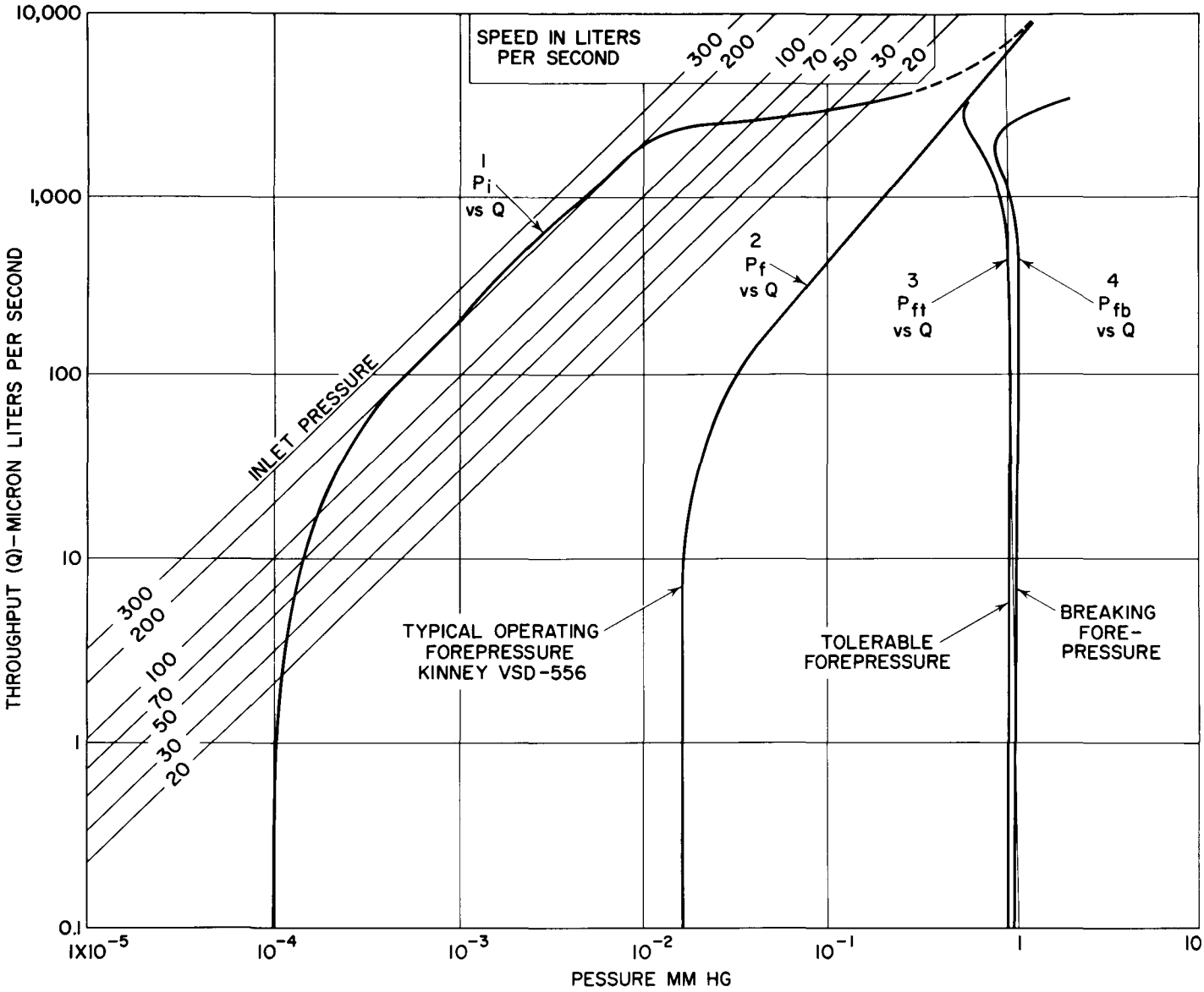


Figure 15. Q-P Curves (Q=gas through put in microniters/sec.; P_1= inlet pressure in mm. Hg; P_f=fore pressure; P_{ft}= tolerable fore pressure — causes 10% increase in P_1; P_{fb}= breaking fore pressure — causes tenfold increase in P_1)

facturing, pressures of 10^{-5} mm Hg and below are achieved by evacuation, and by processing techniques employing bakeout ovens, filament activation, and radio-frequency bombardment. Bakeout ovens are used to drive sorbed and occluded gases from the envelope or vessel being evacuated. Filament activation and radio-frequency bombardment drives the gases from the tube elements. High rates of activation and bombardment should be avoided to minimize pump-oil contamination from excessive outgassing rates. Keeping tube parts clean helps to reduce the gas loads encountered during tube exhaust.

When an evacuated tube cools, gas molecules within the envelope are reabsorbed into the tube parts. At a pressure of approximately 10^{-6} mm Hg this gettering action by the tube may create a reverse pumping action and the pressure in the tube falls below that of the pump. This reverse flow of gas molecules from the pump into the tube is undesirable because of the possibility of contaminating the tube parts. To obviate reverse flow, the tubulation can be tipped, or pinched off, while the tube is hot. A "hot" pinch- or tip-off has a further advantage in that as the tube cools the gettering (absorption) of gas molecules within the envelope reduces the tube pressure.

VAPOR PRESSURE — PUMP MEDIUM

The ultimate pressure which a diffusion pump can

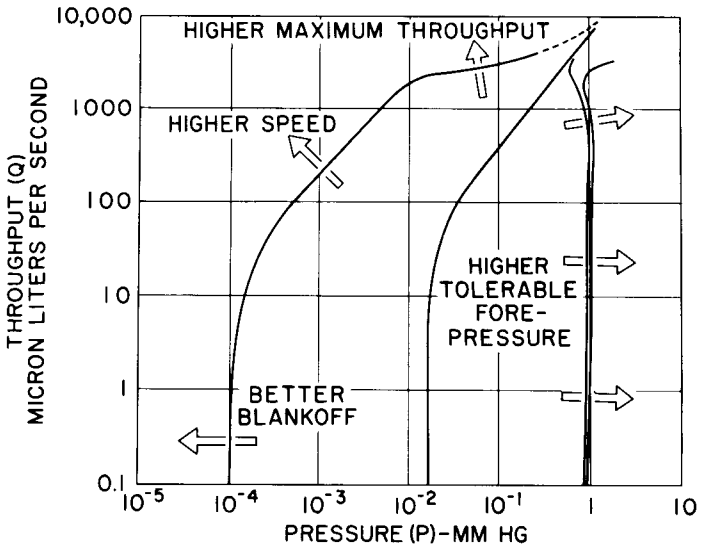


Figure 16. Comparisons With Q-P Curves

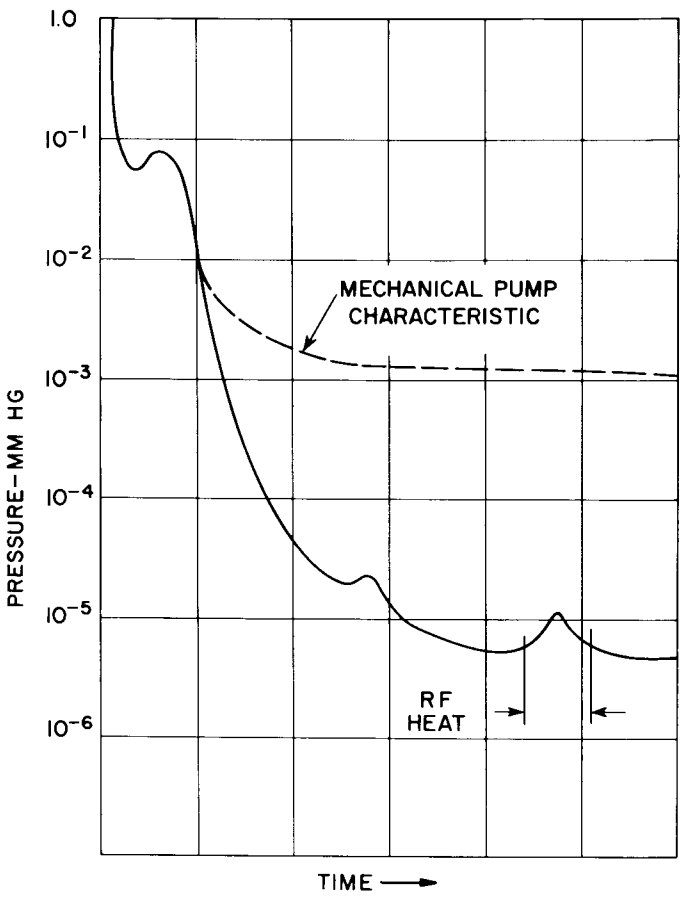


Figure 17. Representative Time-Pressure Curve

produce is often limited by the vapor pressure of the pumping medium. Therefore, it is important that the pumping fluid have the lowest practical vapor pressure. Fig. 18 shows the vapor pressure characteristics of various fluids commonly used in diffusion pumps; water is included for reference. The fluid with the lowest known vapor pressure is Octoil-S. This fluid and Octoil are fine fractions of di-oxtyl sebacate, which is sold by several companies. One of the companies also sells the crude product under the trade name Monoplex. Dow Corning silicone fluids, as previously mentioned, can withstand prolonged atmospheric exposure without deteriorating. Octoils and silicones are expensive, whereas fluids such as Monoplex and Convoil 20 are relatively cheap. With proper operating precautions, the cheaper oils can almost always be satisfactorily used. Butyl phthalate, slightly more expensive than Convoil 20, has been satisfactorily used for booster type applications. Its thermal property has been utilized to obtain shorter cycles for an evaporation process. Compared to the oils, mercury stands out as having a high vapor pressure. In a mercury vacuum system, therefore, a freezing medium such as liquid air or liquid nitrogen must be employed to attain vacuum pressures below 10^{-3} mm Hg. The chemical stability of mercury makes it the only satisfactory medium for systems that exhaust oxygen. Oil-vapor pumps should not be used in systems that exhaust oxygen; if sufficient oxygen mixes with the hot oil an explosion can occur.

VAPOR PRESSURE – METALS

Systems that operate at pressures below 10^{-4} mm Hg should comprise materials having low vapor pressures at the operating temperature. Such materials, particularly when cleaned, will add little or no gas load to the system. Of the common metals, nickel (Fig. 19) and stainless steel have low vapor pressures and are very good for chambers, jig-fixtures, and manifolds in high-vacuum applications. Chrome plating is sometimes beneficial; however, faulty plating can produce deleterious effects, especially in the presence of heat. Gases trapped between the parent metal and the chrome coating can cause an apparent, or virtual, leak. Such a leak is difficult to detect and greatly increases pump-down time.

VAPOR PRESSURE – GASKET MATERIALS

The gasket materials selected for connecting the components of a vacuum system must have vapor pressures commensurate with the vacuum to be attained. The physical properties of the gasket material must also suit the application. Gasket materials suitable for high-vacuum service, listed in order of increasing vapor pressure, include: (1) Teflon, (2) Kel-F, (3) Silicone SR-1060, (4) Convaseal, (5) Silicone SR-1080, (6) Koroseal, (7) Hycar OR-15, (8) Neoprene, (9) pure gum rubber, (10) Tygon, and (11) Garlock 7986.

EVAPOR-ION PUMP

The Evapor-ion pump was developed as an answer to the objectional backstreaming characteristic of all oil-diffusion pumps. The Evapor-ion pump produces a "dry" high vacuum of 10^{-8} mm Hg without any danger of contamination from pumping fluids. Because it can be used without traps or baffles it can operate at maximum speed. Fig. 20 shows a cutaway drawing of the pump. Titanium, evaporated by the filament, chem-

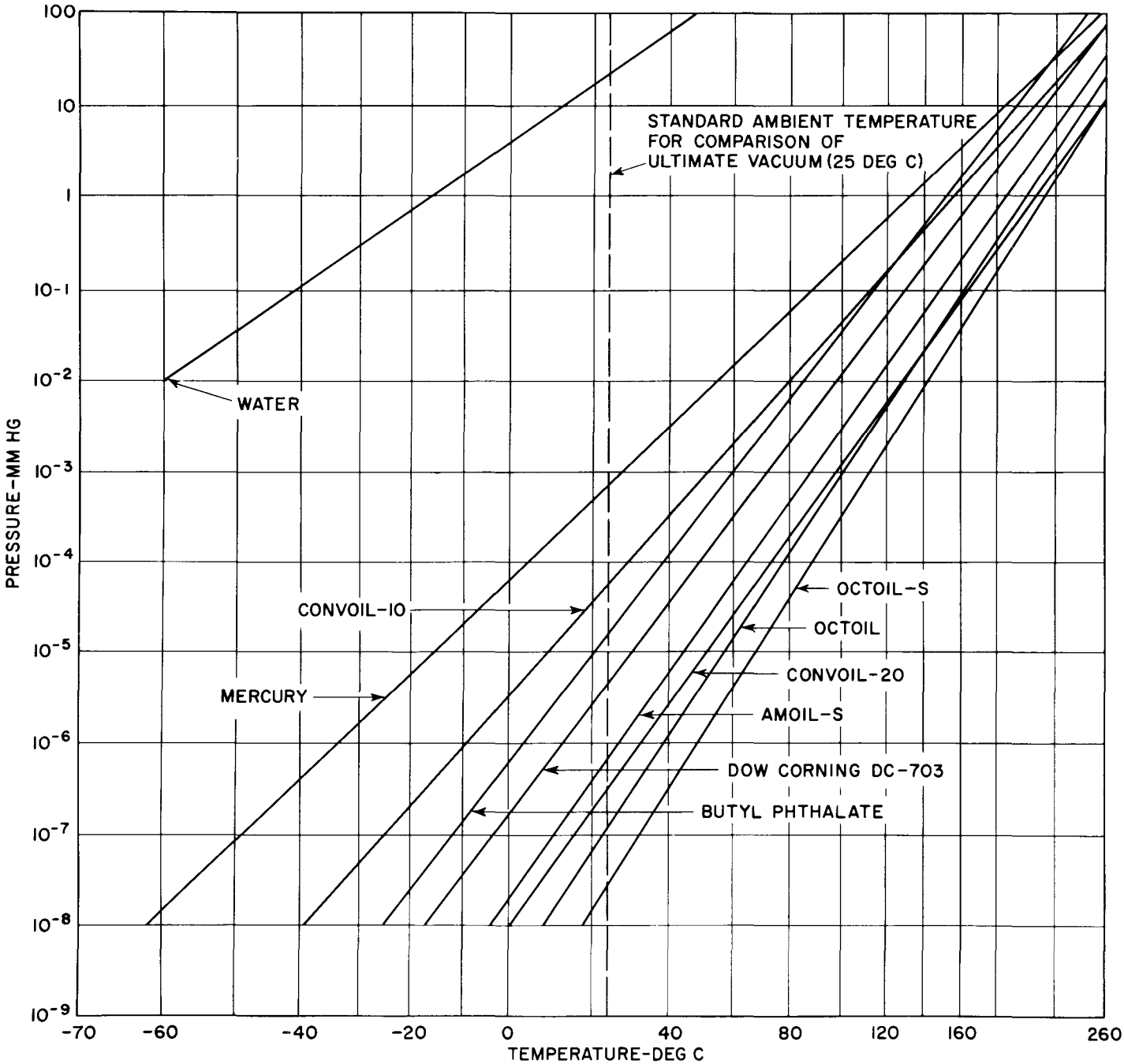


Fig. 18. Pressure-Temperature Characteristics of Pump Fluids

ically combines with active gas molecules and condenses on the water-cooled walls of the pump chamber. Inert gases, ionized by electrons from the filament, are driven by the grid potential into the titanium deposits. As each layer of titanium becomes chemically saturated, a new layer of the metal is deposited. Before the pump will operate, the pump chamber must be "roughed-out" to pressures below 30 microns, the mechanical pump must be valved off, and a primer charge of titanium must be fired by resistance heating to achieve pressures of 1×10^{-4} mm Hg. At this pressure, electrical bombardment and continuous titanium evaporation can be started and pressures of 1×10^{-7} mm Hg can be attained. Pressures of 10^{-8} mm Hg or lower are attained by ion pumping only (i.e., without evaporation of titanium). The use of a conventional oil pump, which can be subsequently valved shut, will eliminate the need for firing of the primer charge of titanium and will reduce pump-down time.

Simplified ionic-type pumps are available under the trade names of "Vac-Ion" and "Ultek."

MERCURY PUMP

A glass mercury pump of the type used on stationary exhaust positions is illustrated in Fig. 21. The mercury contained in the reservoir is vaporized by an elec-

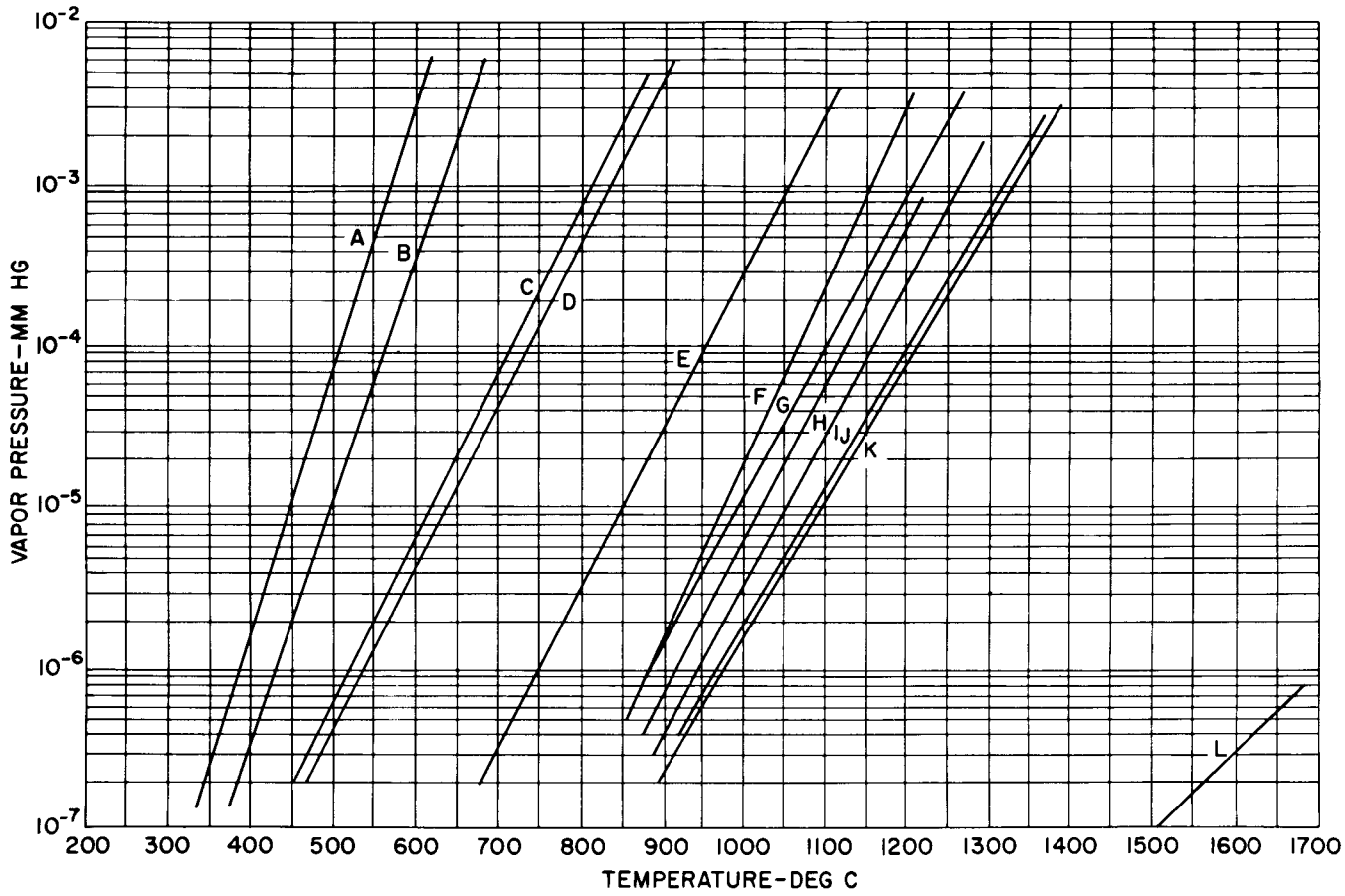


Figure 19. Vapor Pressure Data on Metals and Alloys: A, zinc; B, lead; C, silver; D, BT solder; E, aluminum; F, copper; G, chromium; H, iron; I, Kovar; J, nickel; K, gold; L, molybdenum

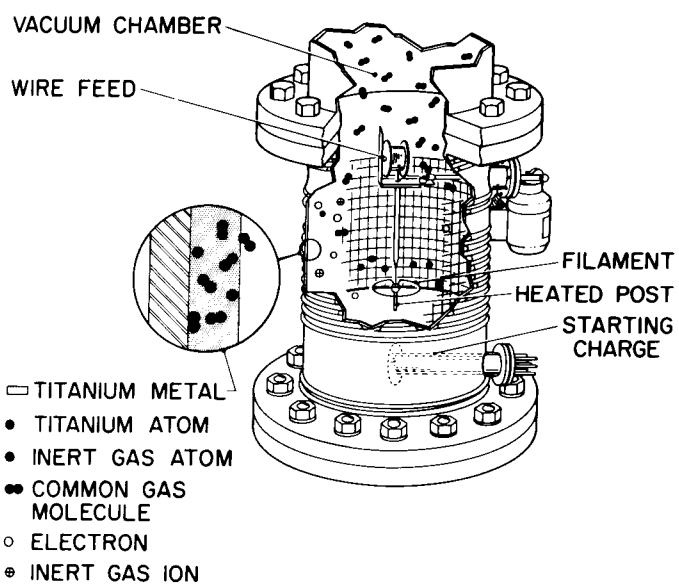


Figure 20. Evapor-Ion Pump

trical heater. Vapor from the reservoir rises and moves at maximum speed through a venturi jet where evacuated gases, entering the sphere from the rear, diffuse into the vapor stream. Water-cooling jackets condense the mercury vapor which returns to the reservoir; the gases continue to stream toward the forepump. In high-vacuum applications the mercury system must always have a cold trap between the pump and the chamber being evacuated. Without the cold trap the ultimate vacuum would be 10^{-3} mm Hg, the vapor pressure of the mercury at room temperature. Mercury pumps can be advantageously used where: (1) large amounts of mercury are involved in the process such as, in the production of mercury-filled tubes; (2) the system must operate against higher fore-pressure than generally possible with oil-diffusion pumps; (3) oxygen is used in the processing as for dynode oxidation to obtain good secondary emitters; and (4) cracking or breakage of the tube or glass manifolds would contaminate the system. Care should be exercised in handling mercury. Its fumes are poisonous, and will attack metals such as zinc, copper, bronze, brass, tin, gold, and silver; and it acts as a catalyst in the hydrolysis of aluminum to aluminum hydroxide. Because of the expense and inconvenience involved in handling the liquid air or liquid nitrogen used in the cold trap, the mercury pump should never be used where an oil-diffusion pump will do.

GAUGES

Pressures within a system can be measured with a variety of gauges. No single gauge, however, is capable of reading the complete range from atmospheric pres-

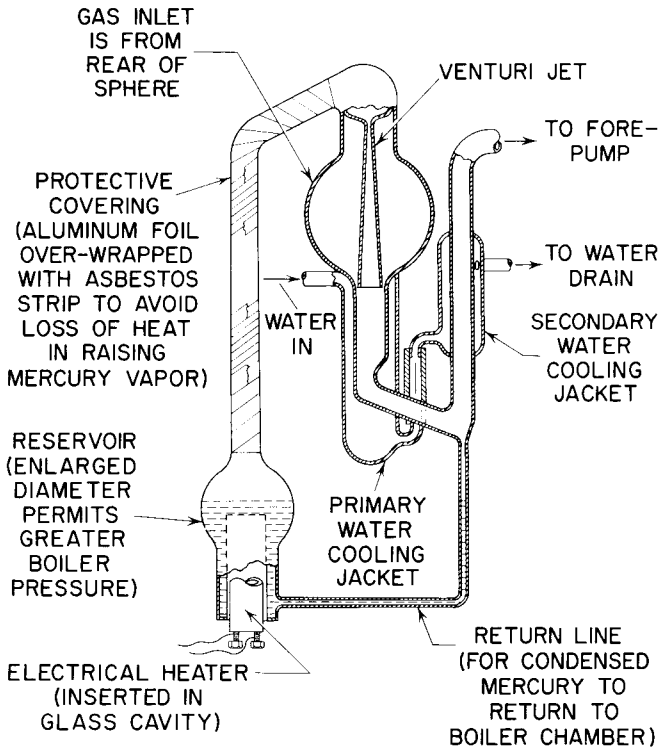

GAS INLET IS FROM REAR OF SPHERE
VENTURI JET
TO FORE-PUMP
PROTECTIVE COVERING (ALUMINUM FOIL OVER-WRAPPED WITH ASBESTOS STRIP TO AVOID LOSS OF HEAT IN RAISING MERCURY VAPOR)
TO WATER DRAIN
WATER IN
SECONDARY WATER COOLING JACKET
RESERVOIR (ENLARGED DIAMETER PERMITS GREATER BOILER PRESSURE)
PRIMARY WATER COOLING JACKET
RETURN LINE (FOR CONDENSED MERCURY TO RETURN TO BOILER CHAMBER)
ELECTRICAL HEATER (INSERTED IN GLASS CAVITY)

Figure 21. Glass Mercury Pump

sure to less than 10^{-4} mm Hg. Fig. 22 gives the range of usefulness of the gauges most commonly used in the tube industry. In practice, the gauge must be selected to fit the application; in some cases two or more gauges will be required to cover the complete range of pressures of a process.

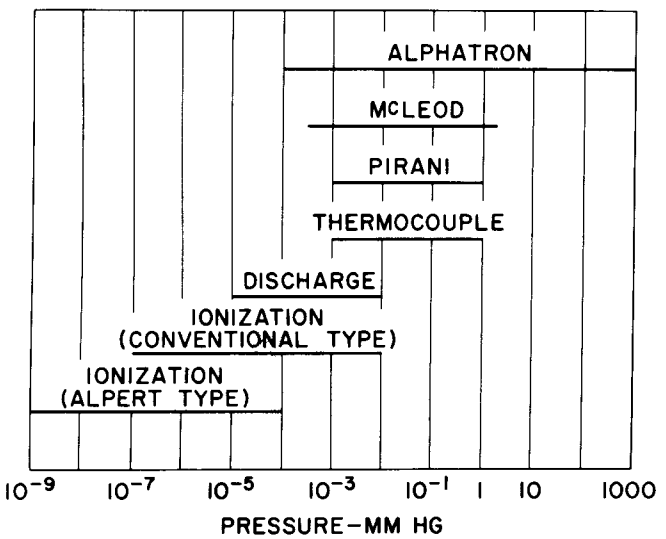


Figure 22. Ranges of Vacuum Gauges

The construction of each of the commonly used gauges is shown in Fig. 23. The McLeod gauge is the simplest and most reliable gauge for direct reading of permanent gases. A volume of gas at unknown pressure within the system is trapped and compressed by mercury into a small capillary tube of known cross-sectional area. The trapped gases are compressed to obtain a pressure difference ΔH readable on a pressure-calibrated scale. The main disadvantage of the McLeod gauge is its inability to measure gases and vapors that condense in the capillary tube. Other disadvantages are the excessive time required for equalizing of system and gauge pressures due to the long, restrictive, connecting tubing; the possibility of glass breakage, particularly if the mercury column is raised rapidly into the capillary; and, in the event of breakage, the dangers to the system or to operating personnel of loose mercury.

In the Alphatron gauge, a small quantity of radium (0.2 milligrams) is contained in a holder surrounded by a grid and a collector. Alpha particles, emitted by the radium, bombard and ionize gas molecules within the envelope. Since the emission is constant, the degree of ionization is directly proportional to the gas pressure. The positive ions formed are driven by the grid voltage to the collector plate; the ion current is amplified and read on a galvanometer directly in pressure units. This instrument is the only one known capable of measuring pressure from atmospheric to 10^{-4} mm Hg. The gauge has no filament to burn out and is less susceptible to contamination than the hot-filament type gauges described later. Objectional features of the Alphatron gauge are the need for frequent calibration of the amplifier to correct for zero drift, the need for recalibration when the gauge head is changed, and the fear instilled in operating personnel because of the presence of the radium. This last objection is not well founded because the radium source is dangerous only when completely exposed. However, plant safety regulations should be checked when a gauge head is to be discarded.

The Pirani and thermocouple gauges are hot-filament types, which operate on the principle that the thermal conductivity of a gas varies with its pressure. In a Pirani tube, pressure changes affect the resistance of the hot filament because of the changes in filament temperature that result from variations in heat conducted from the filament. The filament forms one leg of a Wheatstone bridge; unbalance currents, resulting from the changes in filament resistance, are read on a galvanometer. The thermocouple gauge operates in a manner similar to the Pirani gauge, but a thermo-junction is connected to the filament. Changes in filament temperature cause corresponding changes in the electromotive force (emf) developed at the junction. This emf is read on a galvanometer. The Pirani gauge and the thermocouple gauge are suitable only for approximations because shortly after initial use the filaments become blackened. This contamination affects thermal transfer and gives false readings.

The discharge gauge uses a cold-cathode type of tube in which a high potential is applied to the anode to cause electron emission from the cathode. A magnetic field forces the electrons to travel a long spiral path over which the electrons bombard and ionize gas molecules. The spiral action increases the number of collisions with the gas molecules to provide greater sensitivity and to sustain the glow discharge at reduced pressures, where the electron mean free path is much greater than the distance between electrodes. The positive ions are collected by the anode to produce a current which is in-

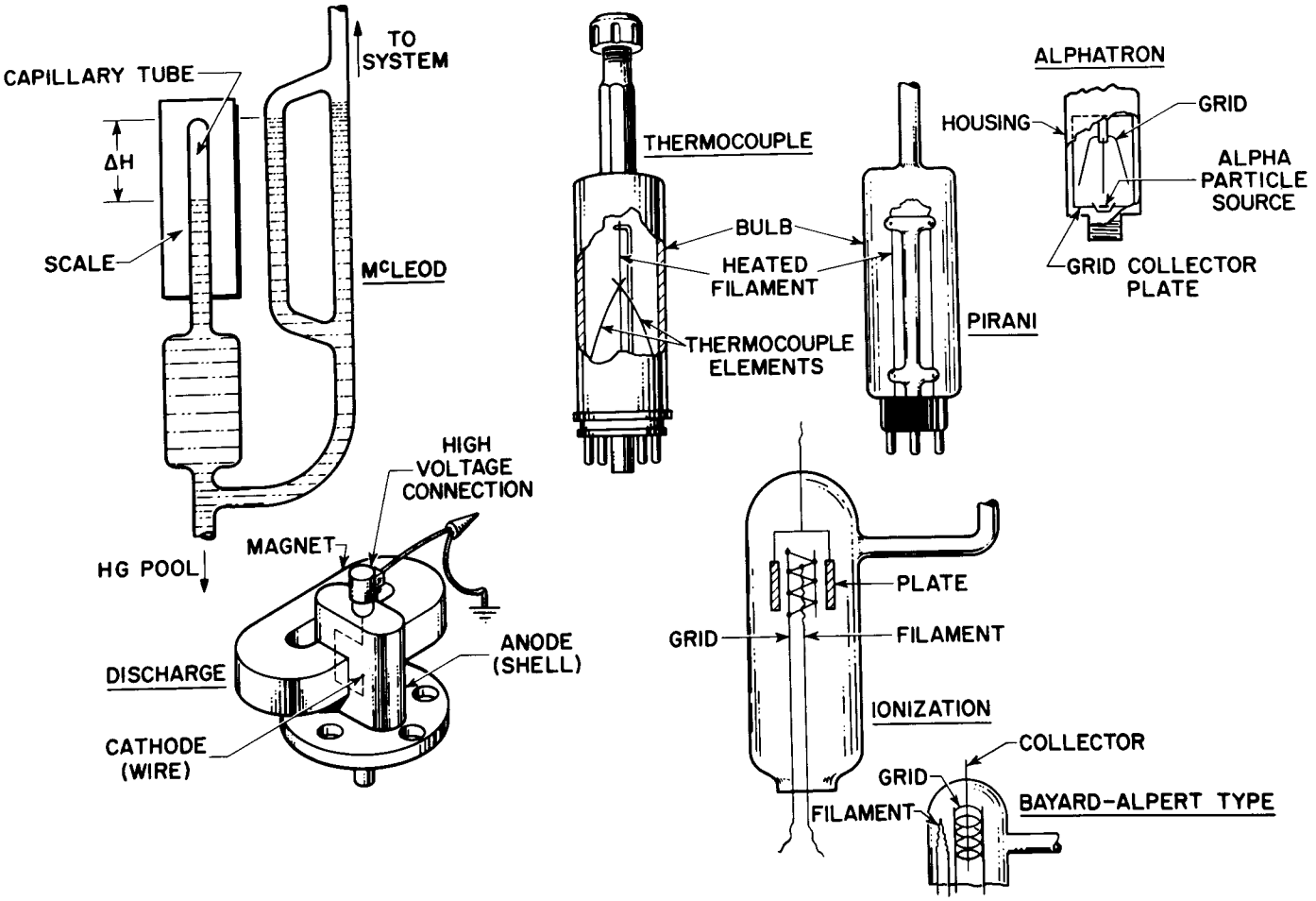


Figure 23. Types of Vacuum Gauges

dicated on a meter. Discharge gauges with an integral magnet, rather than the external horseshoe magnet shown in the illustration, are available and are capable of indicating pressures to 1 x 10^{-7} mm Hg. The gauge with the external magnet is useful for reading pressures to 1 x 10^{-5} mm Hg. As in all cold-cathode-type gauges there is no filament to burn out. The sensing tube is rugged and can be readily washed clean. However, the tube is more easily contaminated than the hot-filament type because it operates at high voltage. Under the influence of the high voltage, backstreaming oil within the tube will break down to form carbon products. Discharge gauges are suitable for reading relative pressures, but are not suitable where exact, reproducible pressure readings are required. An additional disadvantage is that cold-cathode tubes cannot be degassed as readily as hot-filament types and are adversely affected by mercury.

One of the gauges most widely used in the manufacture of electron tubes is the conventional hot-filament ionization gauge tube, such as the RCA 1949. Electrons from the heated filament of this tube are accelerated towards the positive grid. In travel, they collide with and ionize gas molecules. The positive ions, created in the region between the grid and collector plate, are attracted to the negative collector and form an ion current the value of which is indicated on a microammeter. When grid voltage and emission current are constant, the number of positive ions formed is proportional to the pressure. Under these conditions the pressure indications are linear. The RCA 1949, which indicates 100 microamperes for one micron of pressure, provides satisfactory readings at pressures as low as 10^{-7} mm Hg. To ensure accuracy, however, the tube elements must be degassed and the envelope must be clean. To degas the tube elements, the collector plate should be electrically bombarded by high emission current; to clean the envelope, the glass should be torched by a soft brush flame or "baked-out" in an oven. In practice, the torching or baking is often ignored without serious effects. As in all hot-filament gauges, the tube life is limited by filament burnout or shorts. In the RCA type ionization gauge tube at pressures below 10^{-7} mm Hg, electrons strike the grid to produce soft X-rays. The X-rays bombard the large negative plate to release photoelectrons which form a residual current that is constant and independent of pressure. Thus, 10^{-7} mm Hg is the lower limit of the RCA tube. The Bayard-Alpert-type ionization gauge[5] avoids this difficulty. As illustrated in Fig. 23, the gauge structure is inverted so that the plate is a fine wire which offers a small target for X-rays from the grid. This construction change reduces the residual current to about 1 per cent of that in a conventional gauge tube. With the reduction of residual currents, pressure readings down to 10^{-10} mm Hg

can be taken.

SEALS

To attain high vacuum, it is very important that proper seals be used between system components. Several types of seals are shown in Fig. 24. The compression-head design is one of the most widely used methods for sealing a tube to a pumping system. Its seal against atmospheric leaks is effected by squeezing a molded rubber ring around the tubulation. The rubber, compressed by turning a nut, provides an effective seal for tubes having either copper or glass tubulations. Excessive compressive forces on the rubber can cause collapse of the tubing, particularly with partially annealed copper tubulations. Many compression heads are operated within an oven, or close to one, and are, therefore, provided with a water jacket to cool the rubber and lengthen its life.

Where compression-head fittings are unsuitable, the flare-fitting type of connection is often used. As shown in Fig. 24, the tubulation is flared to fit under a nut or sleeve or both so that when the nut is turned the tubulation flare cold-flows on to the mating face of an adaptor connected to the system. The flow of the copper against the mating surfaces forms a vacuum-tight seal suitable for pressures into the ultrahigh vacuum range. With the proper selection of materials and type of thread, this seal can also withstand the highest bake-out temperatures. Undesirable features, especially for high-production operations, are the need to preassemble the nut and sleeve to the tubulation before forming the flare, and the need to salvage and store the nuts and sleeves. For the ultrahigh vacuum range, however, the flared connection is a superior fitting due to the absence of vapor-pressure effects from the rubber. The commercially available gland type of connection is convenient and economical for the quick installation or removal of a gauge. The vacuum-tight seal, effective to pressures of 10^{-5} and possible 10^{-6} mm Hg, is obtained by turning the gland nut to compress the tapered rubber bushing around the gauge tubulation. Flanges with either rubber or metal gaskets may be satisfactorily used to seal components to a system. The cross section of rubber gaskets used can be rectangular, square, or round ("O" ring). The O ring is the most economical because of commercial availability. In use, it is advisable to set the rubber gasket into a grooved flange as shown in Fig. 24. This technique, which provides for sufficient rubber compression to ensure a vacuum-tight seal without excessive distortion of the rubber, permits the gasket to be reused. Use of the groove also minimizes the area of rubber exposed to

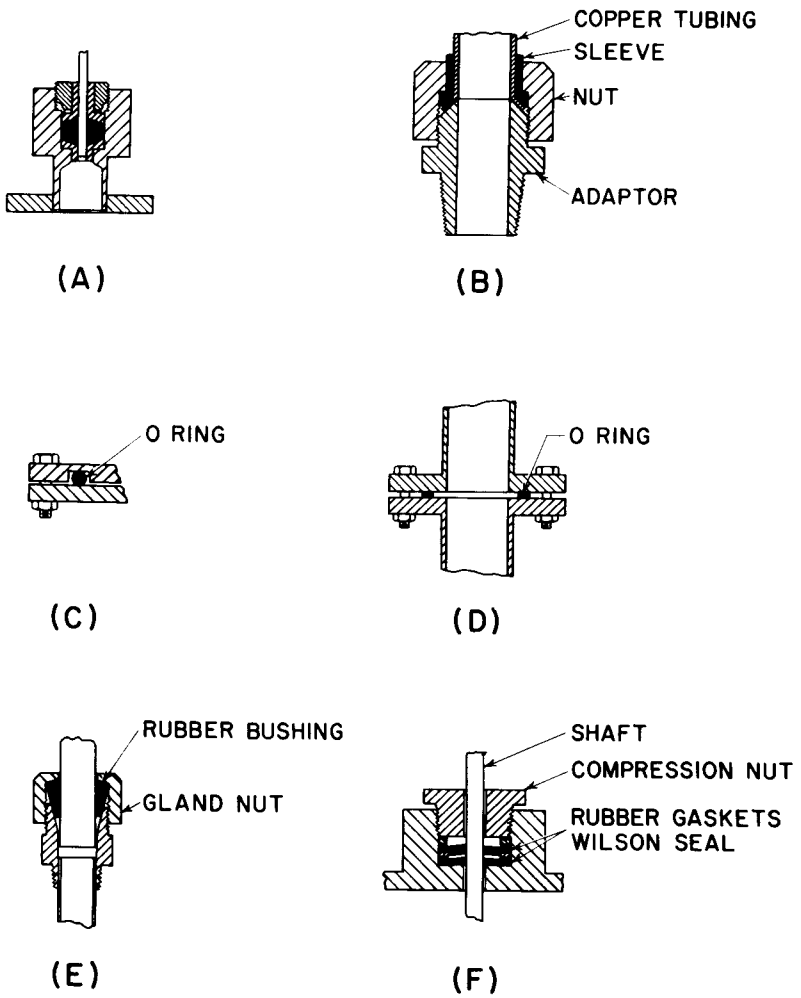


Figure 24. Types of Vacuum Seals: (A) compression head; (B) flare fitting; (C) rubber O ring; (D) copper O ring; (E) gland type; (F) Wilson seal

the system, thereby reducing the effect of vapor pressure from the rubber. Such a seal is adequate for system pressures of 10^{-5} or 10^{-6} mm Hg, provided the system is relatively free of materials having objectional outgassing. The flange surfaces for rubber gasketing should be at least a fine finish (64 microinches, or better) in order to eliminate the possibility of minute leaks.

For system pressures of 10^{-7} mm Hg, or lower, mating flanges must be sealed with an annealed copper or aluminum O ring, preferably copper. The flanges must be a thick section of a hard material such as stainless steel, and their faces should be lapped. When the large forces required to cold-flow the gasket come into play, the thick, hard material will have enough rigidity to prevent concaving of the flange surfaces, and the lapped surfaces will assure vacuum-tight bonding of the gasket to the flange. Metal gaskets should not be reused after mating flanges are parted. An inadequate seal may result in system contamination and tube scrap which is far more costly than the pennies saved by reuse of a gasket. The O ring, although popular and convenient, is certainly not the only metal sealing method. Other sealing techniques, such as the use of a flat gasket and knife-edged flanges[6], may be required.

For applications where mechanical motion must be transmitted into a vacuum chamber, a Wilson seal, or equivalent O ring arrangement, will provide the vacuum-tight seal required of the member passing through the chamber wall. Basically, the Wilson seal consists of one or two rubber disks clamped tightly along their outer edges and having holes smaller than the shaft diameter of the operating member. The restricted hole in the gasket forms an initial seal which is made more secure by atmospheric pressures exerted against it during evacuation. When O rings are used, the rings are spaced to form a double seal; the volume between the rings is usually filled with a high-vacuum lubricant to prevent "drying-out" and deterioration of the dynamic seals. A new and better version of the O ring is the "Quad" ring which, in cross-section, appears as a square with a recess in each of its four sides. Variations of the Wilson seal or O-ring arrangement are applicable for electrical "feed-throughs" into a vacuum chamber.

TRAPS AND BAFFLES

Traps and baffles prevent backstreaming oils and condensible vapors from reaching the vacuum chamber. When a mercury pump is used, mercury vapors, unless removed in a refrigerated trap, would diffuse into the exhausted tube. The trap shown at A, Fig. 25, or some variation of it, is the conventional one used to condense such vapors. Vapors diffusing in the reverse direction must flow to the bottom, or coldest region, of the trap which is immersed in a container of refrigerant, preferably liquid nitrogen (-195 C). It is advisable not to have the container of refrigerant in contact with the trap until the initial stage of pumping has removed the bulk of water vapor from the vacuum chamber. If this precaution is not observed, condensed water vapor at the inlet fringe of the trap may be released later in the exhaust cycle by a warming of this fringe surface. A

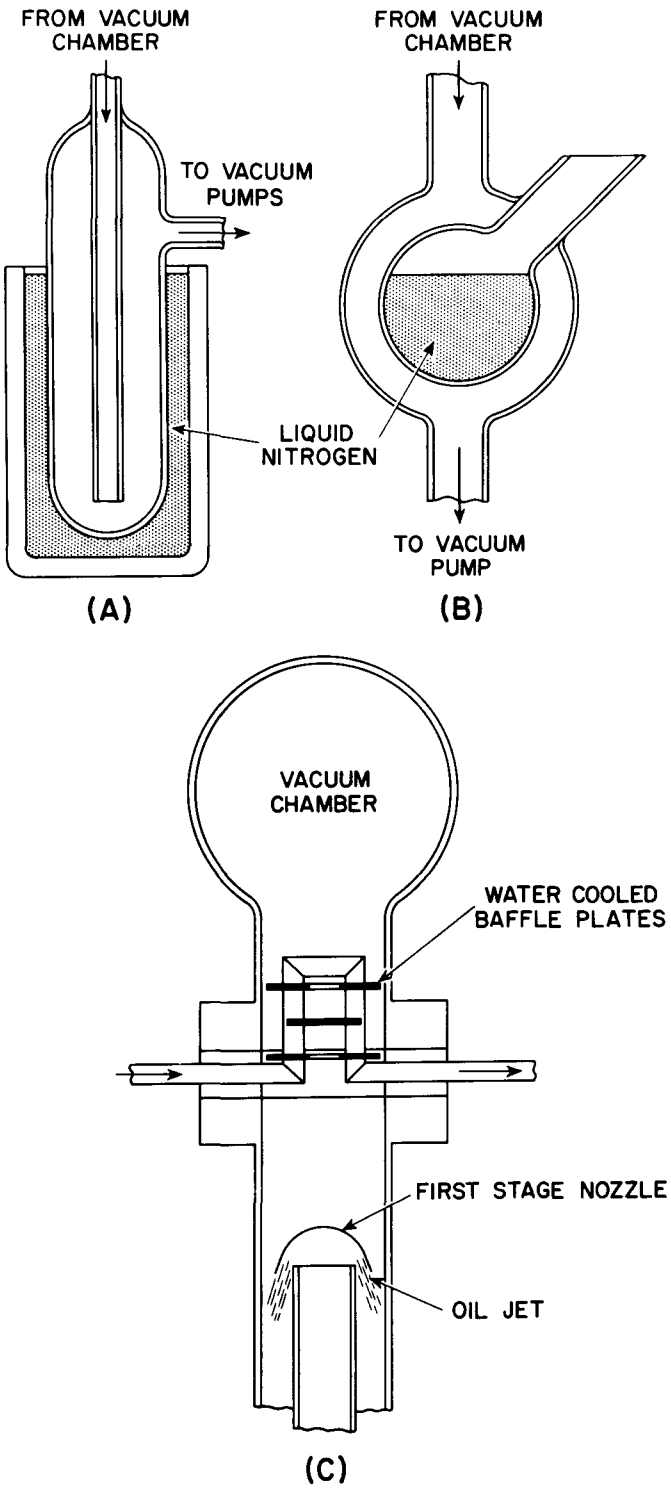


Figure 25. Baffles and Traps: (A) conventional type; (B) all metal trap; (C) water-cooled baffle type

popular metal trap, for use on an all-metal installation for very high vacuum, is shown at B, Fig. 25. The two vacuum-tight spheres are concentrically assembled so that back-diffusing and condensible vapors must impinge on the refrigerated surface of the inner sphere. The coolant can be readily added or removed to conform with the pumping cycle. Because a refrigerant such as

liquid nitrogen is expensive and inconvenient to handle, the trap should not be used unless it is essential to the process. Aside from cold traps, some reduction in backstreaming can be achieved by using baffling above the pump intake. An advanced type of baffle is shown at C, Fig. 25. Water-cooling the baffle plates provides colder surfaces and, therefore, better control of upward migration of vapors. The ideal baffle should be optically opaque so that every molecule traversing the baffle must strike a condenser plate at least once. At best, however, such a baffle can be expected to remove only a major portion (in excess of 50 per cent) of backstreaming vapors. The increased effectiveness evident with cooled baffles makes the use of mechanical refrigeration attractive, particularly in large, fast-pumping systems such as those used in vacuum melting and firing processes. Although the cost of mechanical refrigeration may be high, its advantages often justify its use.

Other traps, not illustrated but worthy of mention, are the charcoal trap and the copper-foil trap. With both of these, the absorbing material is known to have a great affinity for oil vapor. The charcoal trap is simply charcoal contained in a wire-mesh cage fashioned into a variety of patterns. The copper-foil trap is a corrugated copper sheet which, when rolled and inserted into a container, forms a series of long narrow tubes.[7, 8] These traps have the disadvantages of needing periodic cleaning of the gas-sorbing materials, and, because all traps and baffles are restrictions, of reducing pumping speed. They are, however, useful in some applications. In an RCA production system, for example, the copper-foil trap in series with a liquid-nitrogen trap makes possible the attainment of vacuum pressures of the order of 10^{-8} mm Hg.

VALVES

Many types of valves are used to control the direction and rate of flow of gases in a vacuum system or to seal against atmospheric pressures. They vary in size, shape, connection, construction materials, type of seal, and operating mechanism. Vacuum valves should never be confused with pressure valves, for the latter are not designed to prevent slow leak-in along the valve stem. Many commercial valves are of the bellows (or packless) type (A, Fig. 26) with a gasket or "O" ring to provide positive seal of the port passage. Leak-in along the stem is eliminated by vacuum-tight welding of a metallic bellows to the valve plate and body or bonnet. In other commercial valves, the stem is sealed by an "O" ring arrangement. Careful consideration should be given to the selection, the correct installation, and the operation of valves to avoid difficulties. The simplest of valves, used principally in fore-pressure lines or in systems evacuated to the high order of 10^{-5} mm Hg, is the pinch clamp shown at B of Fig. 26. The pinch simply squeezes shut a section of rubber tubing. In addition to the "nutcracker" style shown, a pinch clamp can be adapted for cam, pneumatic, or solenoid operation. A cutoff valve used in glass mercury systems is shown at C, Fig. 26. The flow of gas is effectively cut off when the mercury is forced into the U tube by low-pressure air applied to the top surface of the mercury pool in the reservoir chamber by squeezing the rubber bulb. A glass stopcock controls the air

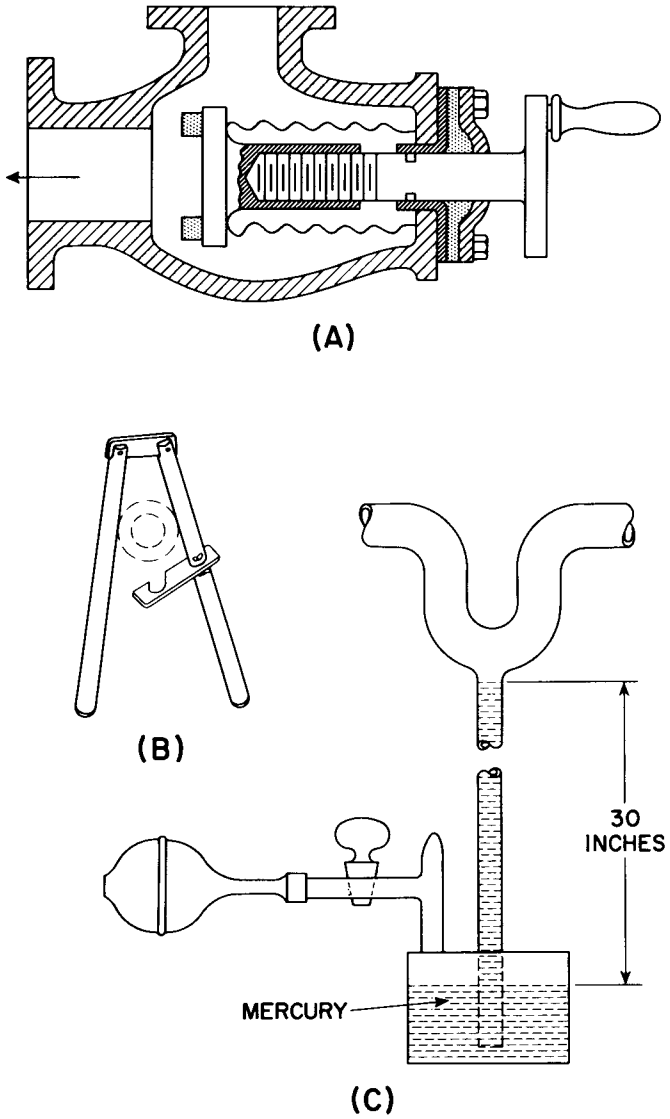


Figure 26. Vacuum Valves: (A) bellows; (B) pinch; (C) mercury

flow to or from the mercury reservoir. Release of air pressure from the reservoir restores the mercury column to the 30-inch level. Cutoff valves of this type, which isolate a chamber or tube from the pumping system, make possible either the introduction of gas for treatment of the tube parts, or the analyses of gases within the chamber. Glass stopcocks for glass systems generally use vacuum grease or mercury to seal against atmospheric leaks. The mercury seals can be troublesome, particularly if mercury vapors within the system cannot be tolerated; the grease seals can be equally objectionable if extreme care is not exercised in the initial application of the sealing grease to the valve plug. The method of applying grease includes not only the complete coating of the plug surface, but also the insertion and oscillation of the plug to assure an even coating over the entire contact area of the plug. Development work by D. Alpert in the field of high vacuum has resulted in a special valve[5] capable of withstanding bakeout temperatures necessary for attainment of a very high vacuum.

The speed of processing, the safety of the operator, and the type and sequence of operations to produce good products without scrap or accidents will dictate the type and quantity of valves employed.

FLOW OF GASES THROUGH VACUUM LINES

The optimum flow of gases through vacuum lines can be achieved by observing the rule of thumb that pipe lines be as short in length and large in diameter as possible and that throttling, due to small valves or port restrictions, should be avoided. The diameter of pipes is by far the most influential factor affecting gas flow. From Knudsen's equation the flow is proportional to the cube of the diameter and inversely to the length of the exhaust lines employed. [9]

A more comprehensive treatment of this subject would be rather lengthy and cannot be covered in this paper. Refs. 10 to 17 cover the subject in detail.

STARTING AND STOPPING PROCEDURE

Much of the servicing of high-vacuum systems can be avoided if the equipment is properly handled and operated. The ordinary operations for starting and stopping a system should follow the procedure outlined below:

To Start:

1. Connect the tube or envelope vacuum tight to the system. Where a bell jar or the equivalent is employed, close chamber and make sure the leak-in valve is closed.
2. Start backer (to achieve the required operating fore-pressure).
3. Make sure diffusion-pump cooling water is flowing at the rate specified by the pump manufacturer.
4. Turn on diffusion pump heater. This can be done at the same time the backer is turned on if the pump fluid will not come up to its vaporizing heat before a limiting fore-pressure is obtained. Whether or not a limiting fore-pressure can be obtained before the fluid reaches vaporizing temperature will depend on the type of pump heater (whether it is of the immersion or pancake type), the capacity of the pump units, volume of the system, and the amount of system outgassing.
5. Raise or apply the trap, if one is used. The procedure for the correct use of the trap was indicated under the heading TRAPS AND BAFFLES.
6. Operate gauges to read system pressure. If the system has only an ionization gauge, be careful to avoid filament burnout. Filament burnout can be avoided by applying only enough filament voltage to provide one tenth of the normal 10 milliamperes grid current. The indicated ion current must then be multiplied by 10 to give approximate system pressure. If the pressure is low enough, the full 10-milliampere grid current can be supplied to the ionization gauge to obtain a true pressure reading.

To Stop:

1. Turn off all gauges.
2. Turn off the diffusion-pump heater and allow fluid to cool sufficiently.
3. If the next evacuation cycle will be delayed for an appreciable period, turn off diffusion-pump cooling water.
4. Turn off backer and leak air into system.

OTHER OPERATING AND DESIGN PRECAUTIONS

In a system containing both rougher and backer pumps (see Fig. 2), the rougher should be valved shut before opening the backer or high-vacuum (hi-vac) side of the exhaust system. The pump heater need not be turned off, but it is important that the hi-vac side be closed before leak-in of air to permit removal of work piece or work load. Incorrect size or position of a leak-in unit can produce deleterious effects on the system. An oversized port could allow a rapid inrush of air which increases the amount of contaminants carried into the chamber evacuated, and the violent inrush could disrupt film and coated surfaces. If the leak-in port is improperly located, for example, if it is in the pump-connecting manifold, inrushing air can sweep over oil laden walls to carry contaminating oil particles onto delicate tube parts or work pieces.

TROUBLE SHOOTING

Difficulties encountered in pumping systems arise from actual or virtual leaks. An actual leak, as the name implies, is a leak due to defective parts, poor workmanship in the fabrication or assembly of the vacuum components, or human failure in the operation of the equipment. The virtual leak is an apparent leak in a system that is actually vacuum tight. The apparent leaks result from the release of gases given off by materials when the pressure within the vacuum system is reduced below the partial vapor pressure of the material. To avoid virtual leaks in high vacuum applications, it is important to use materials having low vapor pressures. An actual leak may be located by isolation. The procedure followed may entail any or all of the following.

1. Reading pressure at various points of the system in a progressive order starting with the mechanical pump only and successively including the sweeps*, diffusion pump, valving, manifolding, and sealing head.
2. Substituting a component known to be vacuum-tight for one that is suspected to be faulty. The pressure gauge is often the trouble spot.
3. Bypassing the suspected trouble spot. This may necessitate removal of a part from the system.
4. Leak-testing the suspected trouble spots. This can be done with conventional leak detecting means such as selected solvent sprays, sparking (for glass systems), or mass spectrometer leak detectors. The leak detection can be carried out on the assembled system or on individual parts removed from the system.

Probable sources of actual or virtual leaks include:

1. Bad Seals. Flange connections, tubing connec-

*A sweep is a curving line of tubing that connects components.

tions, static and dynamic feed-throughs, gauges, bell jars, evacuated tanks, and electron tubes in process.
2. Welds. In all fabricated and commercial parts including brazed and soldered joints.
3. Faulty Components. Mechanical pumps, diffusion pumps, valves, gauge tubes and controls, traps, manifolds, base plates, connectors, and fittings.
4. Operating Difficulties. Water circulation; electrical contacts; bakeout, bombardment, and filament activation processes; tip-off or pinch-off; and use of cold trap.

OTHER PROCESSES

Aside from the evacuation of tubes, sundry other vacuum processes are performed at the RCA Lancaster Plant. Semiconductor evaporation, vacuum brazing, and metal evaporation are covered elsewhere in the book. A brief description of additional processes follows.

Sputtering

Sputtering is a coating process differing from other evaporation techniques in that deposition of metal is achieved by electrical bombardment. The process is employed for the manufacture of fine mesh screens used in targets of image orthicon and vidicon tubes. Schematically, Fig. 27 shows a typical apparatus for sputtering. A negative high potential is applied to the palladium plate causing a glow discharge to form between the cathode and anode, which are the palladium disc and aluminum support respectively. The positive ions formed bombard the palladium disc to dislodge particles which are attracted to the positive plate. A glass master, positioned in the paths of particle travel, becomes coated with a palladium film. In subsequent processing this film is developed into the final mesh. Successful sputtering cannot be achieved with too fine a vacuum because of the need for a substained glow discharge. Pressures of 50 to 80 microns provide optimum vacuum conditions for sputtering.

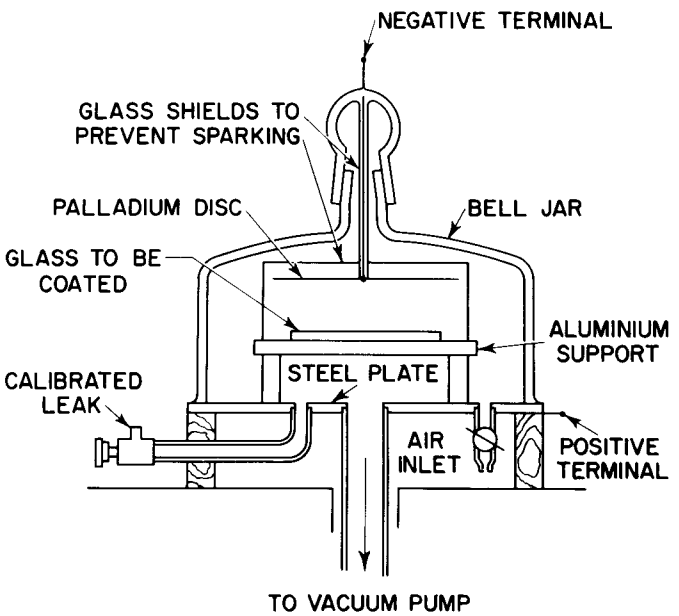


Figure 27. Typical Sputtering Apparatus

Vacuum Firing — AC Heating

Vacuum Firing is performed primarily to remove gases from tube elements and thus minimize contaminating influences in final tube exhaust. The method of heating will vary to suit the part and production required. Ultimate pressures for the varied types of firing operations range from the micron to 10^{-2} mm Hg level. Direct electrical heating of the part is depicted in Fig. 28. Although only one cathode assembly is shown, in reality a gang arrangement is more economically employed.

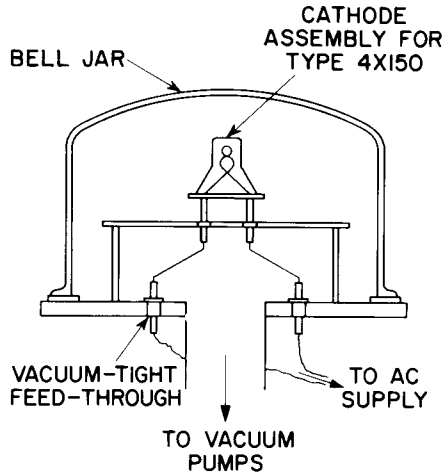


Figure 28. Vacuum Firing (cathode sintering by heating)

Vacuum Firing — RF Induction Heating

Small batch loads of parts which do not lend themselves to ac heating can be readily treated by induction heating as illustrated in Fig. 29. Parts in a metal container are placed in an rf field and heated by induction to drive out occluded gases.

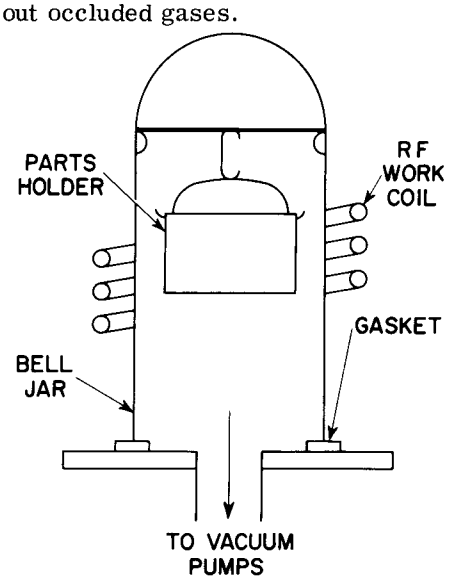


Figure 29. Vacuum Firing (rf induction heating)

Vacuum Firing — Electric Oven

For large work loads, or work pieces, indirect elec-

tric oven heating techniques are employed as shown in Fig. 30. The parts holder is moved into the heat zone by means of the push rod or equivalent arrangement, and returned to the cooling zone prior to unloading. The equipment can be arranged for the work load to move horizontally, as shown, or vertically. It can also be had without a cooling chamber in which case cooling is achieved by removal of the retort from the oven, or removal of the oven from the retort. All vacuum seals are water cooled to prolong the life of the gasket.

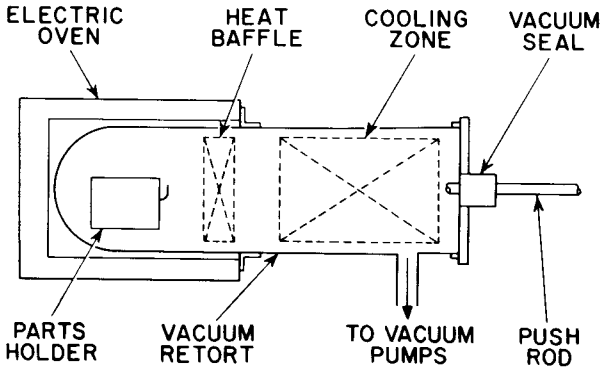


Figure 30. Vacuum Firing (electric oven)

REFERENCES

1. American Vacuum Society, established October 1957, was formerly the Committee on Vacuum Techniques, Inc. started in 1953
2. Committee on Vacuum Techniques, "Recommended Standard Nomenclature in High Vacuum Technology," Vacuum Symposium Transactions, 1954, p. 134
3. Power and Crawley, "Sources, Measurement and Control of Backstreaming in Oil Vapor Vacuum Pumps," Vacuum, Vol. IV, 4, October 1954, pp. 415-437
4. Lawrence, R. B., "High-Vacuum Vapor Pumps, A New Synoptic Presentation of Their Characteristics," Chemical Engineering Progress, Vol. 48, 11 November, 1952, pp. 537-541
5. Alpert, D., "New Developments in the Production and Measurement of Ultra High Vacuum," J. Appl. Physics, 24, 860-76, 1953
6. Hees, G., W. Eaton, and J. Lech, Vacuum Symposium Transactions, CVT, p. 75-79, 1955 "The Knife Edge Vacuum Seal"
7. Alpert, D., "Copper Isolation Trap for Vacuum System," Rev. Scient. Instru., 24, 1004-1005, 1953
8. Alpert, D., "Ultra-High Vacuum Technology," Vacuum Symposium Transactions, CVT, 1954, p. 74
9. Knudsen, M., Ann. d. Physik, 28, 75, 999, 1909
10. Dushman, S., Scientific Foundations of Vacuum Technique, Wiley, New York, 1949
11. Guthrie, A. and R. K. Wakerling, "Vacuum Equipment & Techniques," McGraw-Hill, New York, 1949
12. Van Atta, C. M., The Design of High Vacuum Systems, Kinney Mfg. Co., Boston, 1955
13. Clausing, P., "Uber die Stromung sehr verdunnter gase durch Rohren von belubiger Lange," Ann d. Physik, 12, 961-989, 1932
14. Parratt, L. G., "Vacuum Technique: Speed of Pumping and Molecular Flow," Am. Phys Teacher, 7, 207, 1939
15. Dryer, W. P., Calculating High Vacuum Systems, Hobart Publishing Co., Washington, D. C., 1947
16. Brown, DiNardo, Ching, and Sherwood. "Flow of Gases in Pipes at Low Pressures," Jn. Appl. Physics, 17, 802, 1946
17. Dayton, B. B., "Measurement and Comparisons of Pumping Speeds," Ind. Eng. Chem., 40, 795-803, 1948

BIBLIOGRAPHY

Davy, J. R., Industrial High Vacuum, Pitman, London, 1951

Holland, L., Vacuum Deposition of Thin Films, Wiley, New York, 1955

Strong, J., Procedures in Experimental Physics, Prentice-Hall, New York, 1955

Yarwood, J., High Vacuum Technique, Wiley, New York, 1955

Ho, T. L., "Multiple Nozzle Diffusion Pumps," Rev. Scient. Instru., 3, 133, 1932

Reimann, A. L., "Vacuum Technique," Chapman & Hall, London, 1952

Vacuum Symposium Transactions, CVT, 1953 to date

Vacuum Magazine, all issues to date

Exhaust of Receiving-Type Tubes

C. J. Pearce

Harrison

OBJECTIVES OF THE EXHAUST PROCESS

The evacuation and treatment of the bulb or envelope of a tube and the elements contained therein have two general objectives: first, to remove enough of the gas molecules from the space and from the materials to prevent gas molecules from interfering with the normal electron movement; second, to activate, degas, or otherwise process the surface or the coatings of the elements.

Meeting these objectives in the few seconds that a receiving-type electron tube is undergoing the exhaust process during the high-speed production cycle can be extremely difficult.

CATHODE BREAKDOWN AND BRIGHTNESS TEMPERATURE

Probably the most important step in the exhaust of many tube types is the decomposition of the cathode carbonate coating, a process generally referred to as "breakdown." Breakdown is usually accomplished by raising the cathode temperature to approximately 875C (1150 K brightness temperature) for a portion of the exhaust cycle. During this period the pressure within the tube is maintained below 5×10^{-1} millimeters of mercury (mm Hg) so that the gases produced by the thermal decomposition of the cathode coating, and other gases and vapors given off by the glass, metal, and mica parts, are rapidly pumped out.

Prompt removal of gases is essential to prevent recombination or other undesirable reactions. These reactions would either delay the complete breakdown of the cathode carbonate coating, or "poison" it to some extent. Although further processing may make recovery from poisoning possible, the quality or life of the tube may be affected.[1]

TRACK LIGHTING SEQUENCE

If, in the early stages of the normal exhaust cycle, the cathode temperature is raised too rapidly by application of excessively high heater current, either of two conditions may result. First, the heater base wire may oxidize as a result of its being heated to a temperature too high for the pressure and nature of the atmosphere present in the tube. This oxidation can cause heater-cathode leakage in the finished tube. Second, the heater may burn out. Because of the thermal lag of the heater wire and coating, the bare leg of the heater wire may rise to a temperature where the leg burns out or fuses. In general, the greater the mass of the heater-cathode assembly, the more thermal lag it will have, and the earlier in the exhaust cycle that the "lighting" (heating) of the heater will have to be started. The initial lighting current may vary from 100 to 150 per cent of the rated heater current.

The initial heating prepares the heater for the higher current to follow and releases some gases from the cathode. These gases may protect the cathode by diluting the other gases and vapors present.

The heater current is increased in successive positions (on the sealex machine) to values ranging from 130 to 185 per cent of the rated heater current. The exact percentage of rated heater current depends upon heater, cathode, grid, and plate or shield parameters, the time period of the exhaust cycle, and the temperatures of surrounding elements.

In some unusual cases, the maximum heater current during the lighting process may not be permitted to exceed the normal heater current. One such case is that of a rectifier tube with a close-spaced cathode-anode assembly in which the heater-cathode unit is relatively low in mass and the anode barrel surrounding the cathode is readily heated to a high temperature by the rf-heating coils. On the other hand, a tube with a relatively short, heavy cathode, and a remote anode or shield may require a high lighting current. An extreme case of this type is a tube in which the anode is of open design that attains only a very low temperature, and so radiates little heat to the cathode during the exhaust operation.

After cathode breakdown has been completed, the cathode temperature increases somewhat even though the heater current is held at a fixed value. As a result, it is often desirable to reduce the heater current appreciably during the latter stages of the exhaust process. Once breakdown is complete there is no longer any need to keep the cathode at a high temperature. The cathode should, however, be kept at a somewhat reduced temperature until the end of the exhaust process.

In the case of tube types having (1) lower than normal cathode operating temperatures, (2) transconductance limits at unusually low heater voltages, or (3) a pas-

sive cathode base material, every effort should be made to keep the heater-lighting current, and consequently the cathode temperature, at the minimum values which can provide satisfactory breakdown and activation. For the best over-all result, this condition should be maintained throughout the aging process as well. These precautions greatly reduce the amount of barium which is evaporated from the cathode coating and, therefore, wasted. A low cathode-coating temperature leaves the coating in better condition for activation and meeting abnormal requirements.

There have been observations that as much as 90 per cent of the potentially available free barium is wasted during some normal production exhaust procedures. This loss means that much of the reducing agent in the cathode base material has also been used up and wasted. Such loss of reducing agent probably accounts for some instances of difficulty in attaining activation within a reasonable aging period. At the usual aging temperatures, considerable time may be required to diffuse additional reducing agent through the cathode base metal to react with the coating.

Temperature - Time - Material Effects

Tube temperatures, pressures, and other factors of a similar nature do not occur or change instantly or simultaneously with the application or variation of rf-heating current, application or variation of heater current, connection of the tube to successive vacuum pumps, or with each index position of a sealex machine.

The temperature rise of an indirectly heated cathode lags considerably behind the application of heater lighting current, particularly when the heater is heavily insulated. In the first few positions of lighting, the value of heater current is not indicative of the relative temperature of the cathode. However, for directly heated or filament-type cathodes of very small mass, the filament temperature does change almost immediately with a change in lighting current. In such cases, the filament current is a good indication of the relative filament temperature throughout the exhaust process. Heavy filamentary cathodes, of course, show somewhat more lag than small filaments, and tend to behave more like indirectly heated cathodes. The number and type of supports and the amount and density of the emission coating also affect the heating characteristic.

As cathode temperatures are increased, the time for which the cathode must be held at the higher temperatures can be greatly reduced if the gases released can be pumped out of the tube rapidly enough. After breakdown has been achieved, the barium evaporation rate can be rather high if pressure conditions are favorable, because the evaporation rate is an exponential function of the absolute temperature. A 15-second interval at temperatures in excess of 1250 K can cause the loss of as much as 90 per cent of the free barium produced by reaction of the coating with the reducing agents in the cathode base material. This effect is particularly serious if the cathode base material is passive, or if the cathode coating is somewhat thinner or less dense than usual. When deactivation results from an excessive cathode temperature, the aging time required to restore cathode activity may be greatly increased.

Pressure - Time Relationships

The relationships between pressure and time show lag with either trolley (manual) or sealex (automatic) exhaust of receiving-type electron tubes. For example, equilibrium pressure conditions occur in the tube, exhaust system, and pumps only after a very prolonged exhaust schedule or period. Such a long period is not practical in the case of the exhaust of receiving-type tubes on sealex equipment. So, from the moment the tube to be processed is opened to the vacuum system, until the tube is closed off from the vacuum system by the tube-tipping operation, the pressure relationships that exist in the tube, system, and pump are dynamic and constantly changing. An exhaust system is not to be regarded as a simple hydraulic system in which, at a given instant, the pressure is essentially the same at every point throughout the system. Unfortunately, vacuum or exhaust systems are too frequently so regarded, an error which leads to erroneous conclusions about what conditions actually do exist in the tube or the system during the exhaust cycle.

The pressure throughout the entire exhaust system changes very rapidly when the tube to be exhausted is opened to the system, or when a tube on the system is accidentally broken. Under such circumstances, the air or gas flowing in the system is at relatively high pressure — nearly at atmospheric pressure. At this pressure, the gas flow follows a viscous pattern rather than a molecular pattern. When the pressure drops below about one millimeter of mercury (mm Hg), the gas flow follows a molecular pattern. At pressures of this order, reduced pump efficiencies and system restrictions take their toll and the over-all pumping speed is considerably reduced from the so-called free-air speed. Because all sections of the system are not affected in the same ratio, the pressure differences for different sections of the system can be expected to show wide variations.

As a general rule, the principal impedance in the entire exhaust system is the present exhaust tubulation, the tube used to connect the tube envelope to the system. This limitation exists whether miniature, glass, subminiature, or metal tube types are being considered. It follows that certain pressure differences would exist between the envelope and the system even if the gas flow was constant or was changing at a uniform rate.

In tube processing, however, large bursts of gases are released by sudden rf heating of the metal parts, particularly during the initial applications. In addition, lighting the cathode suddenly releases large amounts of gases, and again pressure changes occur throughout the system. At all these times, the pressure within the tube greatly exceeds the pressures as read in any other portion of the system. The peak pressures, therefore, occur earlier in the tube than indicated by vacuum gauges anywhere in the system. There is evidence which indicates that actual tube pressures under most conditions may be from 12 to 25 times higher than

the pressure read at the exhaust port immediately below the tube.[2] A still higher ratio may exist between tube pressure and the pressure measured by a gauge located nearer the pump.

These results hold true in general both for manual exhaust positions and sealex exhaust machines. Therefore, a pressure reading, such as might be obtained with a Pirani gauge, must never be thought of as indicative of true tube pressure. It indicates only relative tube pressure, and is dependent upon the particular combination of gauge tube and system characteristics. It may show a pressure considerably higher or lower than the pressure actually occurring in the tube.

As a result of the normally very high actual tube pressures caused by the degassing of parts and by the initial cathode-coating breakdown which occurs almost simultaneously or closely following the parts degassing, the time interval that the cathode must be maintained at its breakdown temperature is greatly increased. Because these gases are not pumped out of the tube at an ideal rate as they evolve from the cathode coating, recombination of them with the cathode coating would occur if the temperature of the cathode should be reduced while these gases were still present.

The effect would be essentially the same as that which would be obtained if the cathode was not heated to a sufficiently high temperature during the exhaust cycle; i. e., an incompletely broken down coating. A similar result would also be obtained if the lighting period was too brief, even though the evolved gases were immediately pumped away in each of these two cases.

Altogether, a very neat balance with tube pressure must be obtained among cathode temperature, cathode base material, and coating, as the pressure changes throughout the exhaust cycle. As a result, the lighting time (or the number of lighting positions available) and the tolerable heater lighting currents must be compromised to maintain the optimum cathode temperatures for the various conditions of pressure.

PARTS OUTGASSING AND COLOR TEMPERATURES

In general, degassing of tube elements or parts by rf heating takes place early in the exhaust cycle or schedule for two reasons, whether the processing is being done on the trolley or sealex.

First, the part, by virtue of its material, shape, and size, may have considerable thermal lag and, therefore, heating of the part must be started promptly to allow adequate time for degassing the part.

Second, it is generally preferable to get the bulk of the gases out of the metal parts and, if possible, pumped out of the tube before the cathode and its coating are raised to a temperature high enough to cause the gases to react with them. The highest gas pressure in the tube usually occurs during the second or third position of rf heating of the tube parts.

In some instances, the most readily heated tube part is a shield and not a working element of the tube. In such a case, the working parts must achieve their heating mainly by radiation from the shield. As a result, the active parts heat very slowly to comparatively low temperatures. This problem is particularly emphasized with metal tubes where the metal envelope is heated externally by a sequence of gas flames.

Temperature - Time - Material Effects

The temperatures of parts heated by rf heating coils surrounding the tube may rise slowly as the tube enters a given machine position and fall somewhat more rapidly as the tube leaves the position. Even though more current may be flowing in the following coil, the parts may not necessarily achieve a higher temperature. Depending upon the material of the part, its size and shape, and the size and shape of the rf heating coil, the coupling characteristics between the coil and the part may be such that the temperature of the part remains essentially at the Curie point of that particular metal throughout several positions of the rf-heating cycle.

Because some parts of the tube may never achieve optimum degassing temperature on most tube types, as many positions on the sealex exhaust cycle as possible must be devoted to rf heating. This requirement is especially important for certain carbonized materials which should be heated to temperatures of approximately 850 to 900C for thorough degassing.

Some of the 50-milliampere filament-type miniature tubes, however, may need only two, three, or four rf-heating positions, even at machine speeds in excess of 800 tubes per hour. But during these few positions, the usually bright anode or shield material is heated to its optimum degassing temperature of about 900 to 950 C; the result is adequate degassing for this type of tube.

Parts Processing History

The amount of gases released by tube parts during exhaust treatment and the ease with which the parts are adequately degassed depend largely on the prior processing of the parts, including the manufacture of the basic metal strip and the method of degassing or deoxidizing the metal.

Subsequent cleaning, firing, plating, or coating operations may adversely affect gas content. And finally, the nature of the surface, or the finish, may alter the amount of gas that must be pumped out of the tube during the exhaust process to achieve satisfactory tube performance.

Simultaneous Parts Degassing

With sealex exhaust, simultaneous degassing of all the active elements and the cathode comes about almost by necessity because of the very brief exhaust cycle, 33 to 110 seconds.

With trolley, or manual exhaust, exhaust schedules are almost always considerably longer than for sealex exhaust. The longer schedule can give rise to a condition in which some part is degassing while the temperature of other parts that have already been degassed

has fallen so low that they begin to absorb some of the gases being released elsewhere.

This phenomenon can be impressively demonstrated with a tube having two cathodes that can be independently heated. First, one cathode is "broken down" with the other cathode cold; then, the second is broken down with the first cold. During the breakdown of the second cathode, the vacuum gauge is read; the cathode is then allowed to cool. The first cathode is once again heated while the vacuum gauge is carefully observed. It will be noticed that considerable gas is again driven out of the first cathode. When the second cathode is reheated, the gauge reading shows gas evolution, but not as much as did the first cathode. Each succeeding cycle shows a lower pressure reading, but the experiment clearly indicates that gas freed in one part of a tube may be absorbed by another part and, in turn, passed on to still others.

Metal parts, particularly carbonized ones, can be shown to behave similarly. Any part or surface in a vacuum device that has been well degassed can be regarded as a sponge, wrung dry and ready to absorb any gas that comes its way. Surfaces such as bulb walls and mica, cathode, and other element coatings behave in this manner.

The ideal degassing condition would, therefore, be attained by bringing all portions of the device to their appropriate optimum temperatures simultaneously, and then maintaining these conditions until practically all of the evolved gas is pumped out of the device. However, it is impractical to achieve this completely ideal condition with the exhaust procedures regularly used for receiving-type electron tubes.

Interactions of Cathode Breakdown and Parts Outgassing

Various interactions may result from the two general processes of cathode breakdown and parts outgassing; all are thermal, physical, or chemical reactions, or combinations of these three effects.

The thermal effects are simply changes in temperature of the cathode resulting from radiation of adjacent or surrounding parts which have been heated to high temperatures, or conversely, the rise in temperature of the anode barrel or closely adjacent grids when the cathode is raised to its breakdown temperature. This effect is particularly noticeable in the case of a close-spaced rectifier tube when constant rf heating is used to heat the anode, and the cathode is then raised to breakdown temperature. The increase in anode temperature due to the hot cathode is readily observed, even by the naked eye. It is obvious that the grids in a multi-element mount will be similarly affected by radiation from adjacent hot parts.

Physical effects may occur in two directions. Carbon or hydrocarbons may be driven from carbonized materials onto the cathode coating and micas when their temperature is not high enough to prevent the deposit. This effect takes place early in the exhaust cycle. More gradual heating of the carbonized part, and earlier but gradual heating of the cathode may avoid the difficulty. In addition, improved parts-cleaning procedures may be indicated; parts never can be too clean.

Aluminum may be vaporized from an anode onto the bulb, the micas, or the cathode by improper heating of aluminum-clad materials, an effect which also generally takes place early in the exhaust procedure. Silver, too, can be deposited on bulb walls, micas, and cathode by overheating of a grid that has been made with silver-plated lateral wire to reduce grid emission. This effect usually occurs later in the exhaust cycle after that grid attains sufficient temperature and the pressure within the tube has been reduced.

Barium may be evaporated from the cathode coating and deposited as the metal on mica, grid, or anode surfaces; such deposits will result in leakage, grid emission, or anode reverse emission. But if the metal combines with some of the evolved gases to form an insulating layer upon which more barium can be deposited at some later stage, either during tube processing or tube operation, then erratic behavior such as the blackout or Whippany effect may ensue.[3]

The chemical interactions that can occur during cathode breakdown and parts outgassing are varied and may be complex in nature. Some may take place entirely during the exhaust process, others may be initiated then and become evident only during initial testing; still others may only be revealed much later in the normal operating life of the tube.[4]

If the air initially in the tube is not adequately pumped out by the time the heater is first lighted, the tungsten of the heater wire may react with traces of oxygen present to form oxides which will diffuse throughout the alundum insulation coating of the heater to produce heater-cathode leakage.[4]

If the cathode is heated too much in the presence of certain gases (such as the hydrocarbons) produced by outgassing of metal parts, the cathode coating may be "poisoned". Activation of the cathode may still be obtained, but with some difficulty.

Prolonged heating of some metal parts can lead to partial decomposition of mica supports and release water vapor to react with the cathode. A slightly gray cathode coating usually results. This form of coating is hard and partially sintered, and generally difficult to activate.[5]

Chlorides on any part of the mount structure may also react with the cathode coating to produce a pitted or cracked appearing cathode surface and consequently a very low emission level. Little can be done in this circumstance except to recognize the symptoms immediately. Any remaining suspected mounts should be rinsed in hot water, oven dried at a temperature of 90 C, or somewhat higher, and then exhaust-processed as usual. Most important, the source of chlorides should be tracked down and eliminated without delay.

VACUUM GAUGES

In the exhaust processing of tubes, a wide variety of pressures or vacuum conditions are encountered. Suitable instruments are needed, therefore, for reading or indicating the degree of vacuum in the system or tube, either actually or relatively, during the various stages of the exhaust schedule.

While a pressure in finished tubes in the region of 10^{-5} to 10^{-7} mm Hg is expected, this degree of vacuum is attained in present receiving-type tubes only by the use of getters.

The pressure range covered on sealex-processed tubes may be from 10 to 500 microns (1×10^{-2} to 5×10^{-1} mm Hg) as read on a Pirani gauge, but the actual pressure in the tube is, of course, much higher.

The range required on trolley-processed tubes may be from 1×10^{-5} to 2.5×10^{-1} mm Hg (or 0.01 to 250 microns).

A Pirani gauge using the RCA 1947 fairly well covers the upper pressure range encountered on the sealex, but it should be used with some supplementary gauge, such as the Veeco cold-cathode ion gauge, to cover adequately the pressures encountered on tubes which are trolley exhausted.

(For a detailed description of the many forms of vacuum gauges, the reader is referred to standard reference books on high-vacuum technique, several of which are listed at the end of this chapter. One of these, "Pressure Measurement in Vacuum Systems," by J. H. Leck, deals almost exclusively with vacuum gauges and their proper use. It also treats with the design and calibration of some types of gauges.)

There are two general classes of gauges, those reading in absolute terms and those reading in relative terms. Briefly, an absolute gauge can be calibrated from its physical dimensions, while a relative gauge must be calibrated by comparison with an absolute gauge. The general class designation should not be interpreted as a measure of gauge accuracy; either class, improperly used, can be in error.

Some of the gauges customarily used in the exhaust of receiving-type tubes will be described.

Absolute Gauges

U-Tube Manometer. The simplest form of vacuum gauge used for receiving-type tubes is the U-tube manometer. The liquid usually used in this gauge is mercury, but may be butyl phthalate or some other low-vapor-pressure fluid that does not readily absorb gases or moisture. The difference between the levels of the fluid in the two arms, in millimeters or inches, is equal to the pressure in the system to which it is attached.

The range over which the U-tube manometer is accurate extends from about 20 to 760 millimeters of mercury. Consequently it is used almost exclusively on manual exhaust systems or on sealex machines that produce gas-filled tubes. The manometer measures either the actual gas-filling pressure of the tube, or the regulated pressure of the gas inlet to an auxiliary gas-introduction device. This latter device feeds a measured volume of gas at the regulated pressure into each tube, where it expands to give a large stepdown in pressure.

Another very important use of the U-tube manometer as a vacuum gauge is as the regular leak-detector device incorporated in the third position of the 16-head sealex machine. This special form of manometer has two electrodes sealed into the top of the closed end; they extend slightly more than an inch into the liquid chamber. These electrodes are normally about five millimeters above the level of the mercury under normal vacuum conditions. But when a tube which has a leak in its envelope (such as a cracked seal or stem) indexes into the third position on the sealex, the mercury level in the closed end rises. When the mercury contacts the two electrodes, a solenoid is activated to close the pinch clamp on the port rubber of the third position. The leaking tube is thus prevented from flooding air into the vacuum system and thereby into adjacent tubes when they are connected to the same pump.

Dubrovin Gauge. Another form of manometer, the Dubrovin gauge, makes use of mercury and a float to obtain an expanded scale of 9-to-1 ratio to provide more accurate readings in the region of 1 to 20 mm Hg. This type of gauge is used primarily for gas-filled tubes; within its pressure range, it has a good degree of accuracy.

McLeod Gauge. The McLeod gauge is another device which uses mercury as the measuring fluid. As the mercury level is raised (by any of a variety of means), the gas in the large bulb with the closed capillary tubing on its top is compressed. When the mercury level in the open capillary tubing is brought to the zero line corresponding to the top of the closed capillary, the level of the mercury in the closed capillary tubing indicates, on the calibration scale adjacent to it, the pressure of the permanent gases in the system to which the gauge is attached.

This type of gauge does not, however, accurately indicate the pressure due to readily condensable vapors such as those of water or oil. When this gauge is used (as is frequently done) to check the pressure attained by Kinney or other makes of mechanical vacuum pumps (generally without a cold trap in the line), it must be realized that the true total pressure is much higher than that indicated on the gauge.

The regular single-range McLeod gauge covering 0 to 250 microns, the extended single-range gauge covering 0 to 750 microns, or the multiple-range gauge covering from 0 to 5000 microns in two or more scales, are useful instruments on manual exhaust systems for gas-filled tubes in the specified pressure ranges. The standardizing notices briefly describe the use of the McLeod gauge, give calibration instructions, and discuss its disadvantages.[6]

The McLeod gauge is also frequently used to check out a sealex vacuum system suspected of leaking. It is connected into one of the tube-port rubbers by an adapter, while the machine is stationary, of course; the pressure reading and the time of pump-down to that reading are recorded. These values are then compared with those for another port indexed into that machine position, or with those for the same port indexed into another machine position. Then the gauge is connected in turn to each one of the pumps and manifolds at the rear of the sealex, where readings are taken with the machine stationary, indexing, or with one or more of the connecting rubbers of a "sweep" (tubing) from that manifold clamped off. Comparison of the various readings indicates the location of leaks or blockages.

However, because the McLeod is not a continuously reading gauge and shows the pressure in the system only at the instant the mercury closes off the bulb from the system, it cannot be used to follow rapidly changing pressures. At best, readings can only be taken about every 30 seconds, and even that does not allow time for pumping out the bulb to the new pressure in the system. A Pirani-type gauge is superior for this type of service.

Relative Gauges

Only two forms of relative gauges, the Pirani and the Ionization, are in general use for receiving-type tubes. These tubes have the advantage over previously described absolute gauges, of reading the pressure of readily condensable vapors, as well as of permanent gases. They, therefore, give a better over-all indication of the conditions in the vacuum system or its components, such as the pump.

The various forms of vacuum gauges used in the electron-tube industry, particularly for receiving tubes, should embody the following characteristics: ruggedness, reliability of operation, rapid response, ease of cleaning or replacement, safety from damage by accidental exposure to air, suitable pressure range for application, and reasonable initial and maintenance costs.

Pirani Gauge. The Pirani gauge tube meets most of the foregoing requirements. It has a platinum-alloy filament operating at a temperature below that at which the frequently present oil vapors "crack." Its operation depends upon the temperature-resistance coefficient characteristic of the wire. A constant current flows through the gauge-tube filament as one arm of a Wheatstone Bridge; the temperature of the wire changes as the gas pressure varies in the gauge tube and the system to which it is connected. The resultant changes in the resistance of the filament unbalance the bridge, a condition which is indicated by the meter. The meter can be calibrated in terms of pressure by proper comparison with the readings of a McLeod gauge.[7]

The Pirani gauge tube, the RCA 1947 usually used for receiving-type applications, has inherently rapid response to pressure changes in the system. Within its range and limitations, and with suitable operating care, it provides a valuable way of continuously monitoring pressure changes in a vacuum system. It does so almost instantaneously as compared with other types of gauges, or with some similar gauges of other manufacture. However, it is not calibrated directly in pressure units but this limitation, in general, does not affect its usefulness. A calibration curve for air is available, however, over the range from about 5 to 500 microns (see Fig. 1).

The Pirani gauge provides a fairly good means of determining the relative pressure changes in the tubes being exhausted. It also reveals the operating condition of the vacuum pumps and other components of the exhaust system.

During the latter part of the exhaust process, the cathode breakdown and parts degassing have been completed, and the pressure in the tube gradually decreases. Generally, in the position just before the tip-off of the tube, the getter is flashed. Although the Pirani gauge may not indicate it, because its sensitivity is low in this region, the actual tube pressure now becomes much less than the system pressure by a factor of possibly one or two magnitudes. (This fact again demonstrates the dynamic nature of exhaust systems, particularly of high-speed automatic units, and emphasizes that pressures, low and high alike, are not attained instantaneously everywhere throughout a given system.)[8]

In any analysis of an exhaust process or problem, it is essential that the pressure-vs-time relationship be taken into consideration; it is an important factor.

An averaged plot of Pirani gauge readings on several tubes (even just the current values, not converted to approximate pressures) against the position number on the sealex, gives a rough picture of the pressure-time relationship that normally exists during the exhaust cycle. Other pertinent data should be noted on the plot such as date, machine number, machine speed, type of rf coils, rf current in the various coils, heater currents, pump complement, and manifolding arrangement. Such data, even if only in tabular form, may be a very useful reference at a later date if some serious trouble arises.

Too often this procedure has been overlooked, and time and money are wasted and production lost because such data were not available for comparison. Such information frequently can resolve production difficulties and really expedite their cure under many conditions. Similarly, the tabulated data alone, or the curve plotted from it, can be very helpful in setting up the sealex track-lighting and rf-heating schedule for a new tube type. If the new type is roughly comparable to the prototype, then the Pirani readings in each machine position from No. 2 to No. 14 should closely approximate those for the prototype.

However, if the new type has an anode or shield of radically different material or size, or has a considerably different size of cathode, a set of Pirani readings should be taken as soon as the schedule for rf heating and track lighting appears to be shaping up in the right region.

The plotted curve should then be compared with some satisfactorily operating type. The curve, or envelope,

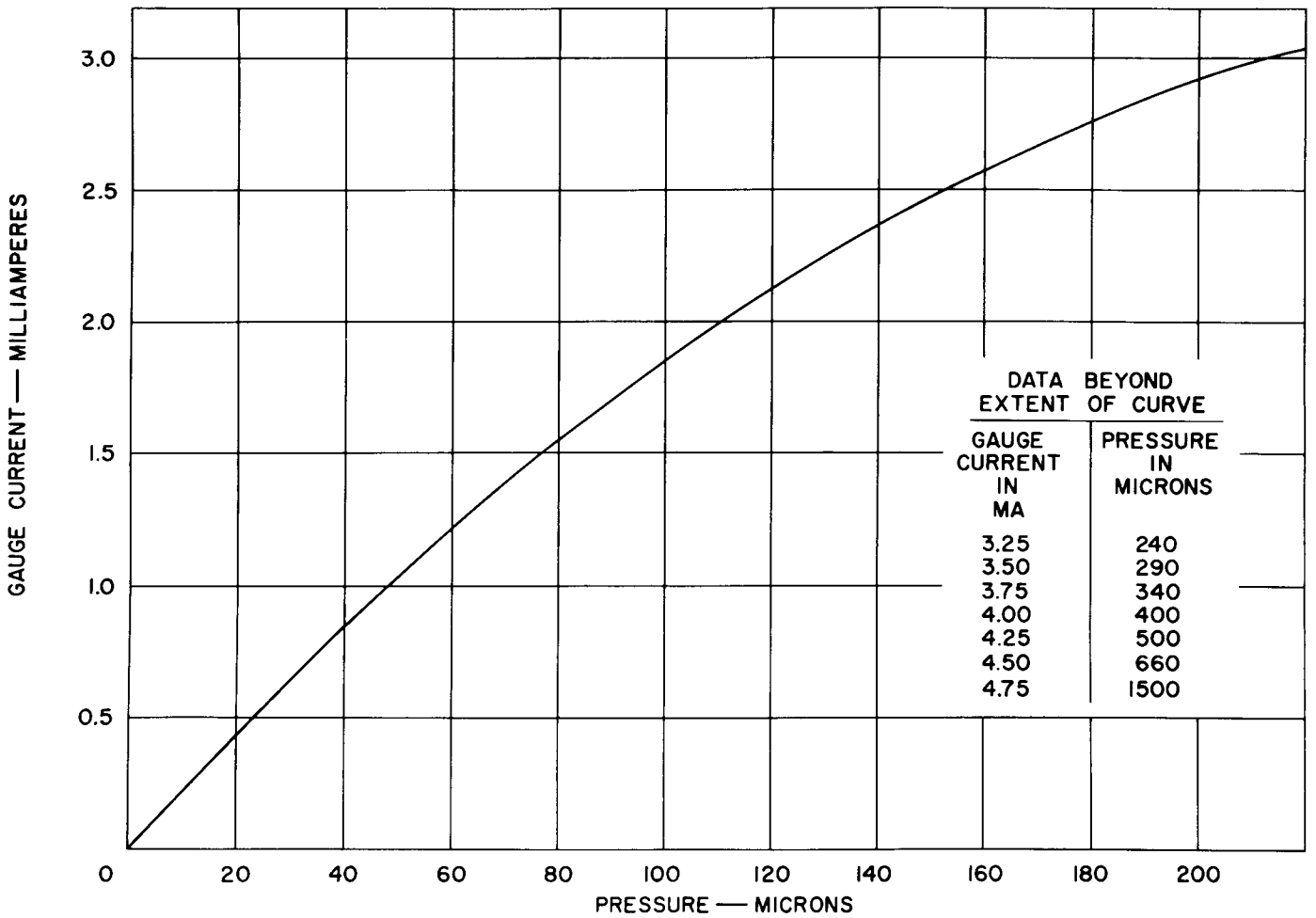


Figure 1. Pirani Gauge Calibration Curve for Dry Air (RCA 1947).

should be similar in shape and, although its values for positions No. 5 to No. 11 may be higher or lower than in the comparison curve, it is important that positions No. 12 to No. 14 show similar or lower values.

The deviations in values for positions No. 5 to No. 11 will indicate whether revisions in rf heating and track lighting are in order. An improved schedule should then be tried and the results plotted again. The operation should be repeated until both a satisfactory shape of envelope and good quality tubes are obtained.

Although McLeod gauges have been generally used to check the condition of the mechanical vacuum pumps on receiving-type sealex machines, the recent trend has been to install a Pirani tube at each pump. This procedure permits continuous dynamic observance of the pump performance, especially under machine-loaded conditions. CAUTION: The Pirani gauge tube used should be an all-metal model, so that any accidental shock to it will not break a glass bulb or enclosure and allow glass to get into the inlet pipe of the Kinney pump where it can cause serious and costly damage to the pump.

Ionization Gauge. The use of the hot-cathode form of Ion gauge (this abbreviated designation is frequently used in preference to "Ionization") can result in the "cracking" or decomposition of certain vapors and thereby give rise to inaccuracies, which usually are not significant for receiving-tube exhaust systems.

The RCA 1949 Ion gauge tube is frequently used for indicating pressures in vacuum systems from 1×10^{-3} mm Hg, down to 1×10^{-6}, or even into the 10^{-7} mm Hg region with an appropriate power supply unit. Pressures of this order of magnitude are attained on manual exhaust systems only in the various laboratories. However, pressures of this order, or lower, are regularly attained on microwave-tube exhaust systems both in laboratory and production use.

This type of gauge tube has a hot tungsten filament as an electron source and is very susceptible to damage by admission of any air, even at relatively low pressures.

A cold-cathode type of Ion gauge, known as the Penning or Philips gauge, depends upon a high voltage and an intense magnetic field to ionize the gases in the gauge tube; it provides a current reading which varies with

pressure. The range of a commercial form of this instrument extends from 1×10^{-2} mm Hg to slightly less than 1×10^{-5} mm Hg. This range is a useful one that adequately covers the usual laboratory systems for receiving-type tubes. Its primary disadvantage, however, is that it functions as an ion-pump during operation, it collects gases from the system and tubes being processed so that eventually its surfaces become contaminated, its behavior becomes erratic and cleaning is necessary to restore normal operation.

A compromise is frequently made in the use of this gauge by operating it only as a monitoring device at certain intervals, instead of operating it constantly. When this procedure is followed, a Pirani gauge is used continuously to indicate pressure changes within its range.

FORMS OF RECEIVING-TYPE EXHAUST EQUIPMENT

Manual Exhaust (Trolley)

The term "trolley" is used to refer to manual exhaust equipment on which electron tubes are evacuated and treated. Its usage dates back to the early days of incandescent lamp manufacture. The term is used similarly in many other countries.

Before the days of automatic sealing and exhaust machines, all incandescent lamps were sealed by hand, or on 6- or 8-head sealing machines. After sealing, the lamps were exhausted on manually operated units arranged in long rows. Each unit held 6, 8, or 10 lamps, depending on their bulb size.

The exhaust operator sat in a car having empty racks on one side to receive the lamps that had just been tipped off. On the other side of the car were racks of sealed lamps ready to be exhausted. Starting at one end of the row, the operator removed a load of lamps from the exhaust unit and placed them in a rack on the car. The operator then took a fresh load of lamps from the car, put them in the exhaust unit, and started their evacuation and bake-out. The operator then moved on to the next unit and repeated the operation. When he reached the end of the row, he moved the car back to the starting point, and started all over again.

The car used in this operation ran on miniature trolley-car wheels and rails. The term "trolley," still universally used for this method of exhaust, came from the use of the wheels and rails, and not from any form of overhead power or feeder.

During the manual exhaust of electron tubes, there are more opportunities to approach ideal conditions than on the usual automatic equipment. For example, each tube generally is processed individually, so far as rf heating of the parts, heater lighting, and getter flashing are concerned. Thus, temperatures, lighting current, and the amount of getter flash, as well as the time intervals, can be carefully controlled or varied as desired to achieve product uniformity.

With adequate equipment and care on the part of the operator, a product almost as uniform as that taken from automatic-type equipment can be obtained by using manual exhaust. It can also be used to simulate the exhaust schedule of sealex machines at normal speeds. This procedure can be particularly useful in the transition from laboratory-processed developmental tubes to factory-processed sealex tubes. Such a procedure, carried out early in the development of a tube, can point out production difficulties ahead, and allow time for re-evaluation of design. The value of this procedure must not be underestimated.

Manual Exhaust Schedule for Vacuum-Type Tubes

The standardizing notices give several trolley exhaust schedules or procedures for receiving-type vacuum tubes. Basically, these schedules include the following four steps:

(1) Bake out, usually for 30 minutes at 400 C.
(2) RF heating and heater lighting; this step is tailored to the tube type and its requirements; it may take from 80 to 220 seconds, or more, or may be chosen to simulate sealex exhaust.
(3) Tip off the tube.
(4) Flash the getter.

Heater lighting, as a rule, is continuous and generally does not exceed 150 per cent of normal heater current. If this percentage is exceeded, it is done only briefly and then reduced.

In some instances, the heater current is brought up slowly so that it does not exceed a chosen value at any time. The cathode-heating rate is considerably slowed by this method, which may be desirable if the anode or other metal parts are being slowly heated by rf coils and are slowly giving up their gases. Such a procedure may also be followed to prevent bowing of the cathode by matching its expansion more closely with that of the rest of the mount-cage assembly.

Another method of heater lighting is sometimes called "snap-on." In this case, the heater voltage is already preset and is switched on instead of being gradually increased by a variable control.

Initially, of course, this method applies heater current well above the value finally obtained at the completion of breakdown. However, this method simplifies the handling of controls and is advantageous for a lone operator. The thermal lag of the heater and cathode moderates their rate of heating to generally satisfactory conditions.

In unusual cases, the heater lighting may need to precede the rf heating of the metal parts. This pattern is indicated in some cases where the heater-and-cathode assembly has considerable thermal lag and the other mount elements are brought up to temperature quite rapidly by the rf heating.

Because the pressure in the system due to the gases released from the tube is under continuous observation on an appropriate type of vacuum gauge, the period of time required for complete breakdown of the cathode

coating is readily observed. Most trolley schedules are based on such vacuum-gauge observations, not only for heater-lighting current and time, but also for temperature of rf-heated parts and time.

It is generally true that cathode breakdown and outgassing of parts are more thoroughly and completely done on trolley exhaust than on sealex machines for the following reasons:

1. The tube is given a long bake-out, which degasses the bulb and drives out volatile materials from other parts.
2. The parts are usually rf-heated for a long period of time under low-pressure conditions.
3. The cathode coating is broken down at a low maximum temperature applied over a long time.
4. The tube is pumped out during the tip-off operation.
5. The getter is usually post-flashed.

When difficulty is experienced with gas shrinkage on sealex-processed tubes, lots of 10 or more tubes are frequently processed on a trolley to determine the source of the trouble. This procedure is quick and simple and expedites the solution of the problem; more frequent use should be made of this procedure.

Manual Exhaust for Gas-Filled Tubes

Certain tubes, referred to as "gas tubes," are filled with a rare gas or gas mixture to a pressure of from 50 microns to more than 50 mm Hg. The gases most frequently used are argon and xenon, or mixtures of neon and argon, or helium and argon. The amount of argon in the mixtures may vary from 0.5 to 20 per cent. Often this percentage must be closely controlled because small variations may make considerable difference in the breakdown voltage or in the operating characteristics of gas tubes.

These gases, as obtained from the vendor, are classed as spectroscopically pure. Because every effort must be made to keep these gases pure until they are used in tubes, extreme care must be exercised to avoid contamination in the transfer of the gas from its container to the tube.

Gas tubes are first evacuated and processed similarly to vacuum-type tubes. The pump is then closed off from the system. Because the gas sweeps through the system in the filling operation, the tubes and the system must be free of leaks or contaminating residue.

Two precautions help to maintain a high degree of gas purity. First, there is a cold trap in the system through which the gas is introduced. Second, in gas-type tubes, the getters are usually flashed before the gas mixture is introduced. In certain instances the getter may be post-flashed, but this practice is not always easy to do or desirable, particularly if the pressure of the gas filling is in the region of several millimeters. Getters flashed under relatively high-pressure conditions give dark deposits and, although these deposits are very active, their appearance is generally disliked.

A manual exhaust position for gas tubes is usually arranged so that the manifold is baked out with each load of tubes. This method prevents contaminants from condensing in the manifold. The manifold and its connecting tubing have relatively small diameters so that the volume of the system is kept to a minimum to prevent waste of gas at each filling.

Gas tubes are usually evacuated in batches of 10 or 20 at a time instead of individually. After this process is completed, the tubes are cooled, the gas is introduced, and the tubes are tipped off.

Automatic Exhaust (Sealex)

The term "sealex" comes from a contraction of "seal" (the operation of fusing the glass bulb to the glass stem) and "exhaust." In a sealex machine, these two operations are combined in a single machine. The sealex machine was developed originally by General Electric for the processing of lamps. The machine made it practical to transfer lamps while still hot from the sealing operation immediately to an exhaust position. Earlier separate machines required reheating of the lamp assembly during the first exhaust stages. The basic lamp machine has since been modified in many respects to make it a highly efficient unit for the production of electron tubes.

The sealex machine is mechanically indexed so that every operation throughout the sealing and exhaust cycles takes place at predetermined and regular intervals. It should be noted that although there are variations in the dimensions of individual stems and bulbs and consequent variation in the rate of heating parts during the sealing operation, and although slight variations occur during exhaust because of small deviations in placement of the tube in each rf-heating coil, as well as variations in cathode-coating weight and heater location, all these variations from tube to tube are usually so small that the sealex is capable of producing a uniformly high quality product.

Sealex-Exhaust Schedules for Glass-Type Vacuum Tubes

The standardizing notices give numerous sealex-exhaust schedules for receiving-type electron tubes. The receiving-tube sealex has 16 positions in the exhaust section. As for manual exhaust, the sealex schedules are tailored to the tube type and its requirements, but, in general, rf heating of parts and heater lighting start at position No. 5, the getter is degassed on position No. 13 and flashed on position No. 14, and the tube is tipped off on position No. 15. Exhaust speeds may run from about 400 tubes per hour to 960, or more.

In the case of 50-milliampere filament-type tubes, the dissipation is small by comparison with indirectly heated cathode types. Therefore, a plain or bright metal anode or shield is used. Because this part degasses quickly, and particularly because there is little cathode coating to release gas simultaneously, only two or three rf-heating positions are needed. As a result, several machine positions are available for pumping

out the tube prior to lighting the filament near the end of the exhaust cycle. Any interactions between the gases coming from the metal parts and the hot filament are minimized. As a general rule, two positions of filament lighting are adequate for this class of tube. Because a fine-filament type is readily damaged by over-lighting, it may be desirable to tend toward under-lighting.

The trend is to use a high heater-supply voltage. This practice minimizes lighting-current variations caused by variations in contact resistance or heater resistance. Because an increased value of resistance is required in series with the tube heater, the initial surge current is reduced in each position as lighting takes place. In the case of the filamentary miniature tubes, an unusually high filament supply voltage is used to reduce the incidence of "open" filaments. Occasionally, a filament tab will be "insulated" from the filament by traces of emission coating or oxide on the base wire. The high voltage breaks down this barrier and produces a good weld.

Bulb Heating During the Exhaust Cycle. Although the tube bulb is still hot from the sealing operation when the tube is loaded into the exhaust port, it is advisable to heat the bulb further. Usually, this heating is done preceding the initial rf-heating coil on positions No. 3 and No. 4.[9] Additional burners on one or more positions (such as positions No. 7 or No. 8 and positions No. 11 or No. 12) to heat at least 12 millimeters of the exhaust tubulation and dome of the bulb are often helpful and are generally good insurance against low emission caused by gray-cathode coating. On occasion, some trace of hydrocarbon contaminant may condense in the relatively cool dome and adjacent exhaust tubulation. If these parts are not heated enough to drive the contaminant beyond the point of tipping off, it remains sealed in the tube to cause trouble.

Glow Discharge to Clean Internal Surfaces. A high-voltage glow discharge in a low-pressure atmosphere has long been used to clean metal or glass surfaces by ion bombardment, prior to coating the part by metallic evaporation.

This principle can be used to advantage in the exhaust of electron tubes, particularly in the case of miniature tubes or similar types which are exhausted at high machine speeds.

A button or similar shaped electrode is mounted at position No. 4 of the exhaust cycle on a Transite or Mycalex support just a few millimeters above the tube pins and connected to the output terminal of an ac-operated spark coil, capable of producing a spark about one-inch long.

When the tube enters position No. 4 during exhaust, the pressure within the tube envelopes is in the region of a few hundred microns. The electrode causes an initial glow discharge which may be somewhat pink or purplish in color, but which gradually changes to a blue or a pale blue if air is not leaking into the tube. The ions in the glow discharge bombard the internal glass bulb wall, the metal elements, and the micas and knock off gas molecules from their surfaces. This process speeds up the removal of much surface gas.

A slow leak in the tube is visually indicated if the pink or purplish glow persists. Such a small leak will not actuate the usual U-tube leak detector in position No. 3 of a 16-head sealex. The slow discharge, therefore, acts both as a sensitive leak indicator and as a means of loosening gas molecules, particles, and even lint.

Sealex Exhaust of Metal-Type Vacuum Tubes

The standardizing notices also give sealex-exhaust schedules for metal-type vacuum tubes. The important difference between the schedules for glass tubes and for metal tubes is that with metal tubes it is not feasible to heat the tube parts by means of rf coils. Instead, the parts are heated by radiation from the metal shell which is heated locally by gas-air-oxygen flames arranged to heat the mount-cage zone. This heating is done in exhaust positions No. 3 to No. 7, inclusive. In positions No. 11 and No. 13, the top of the shell of some tube types is heated by gas-air-oxygen flames to flash a pellet-type getter affixed to the inside of the shell dome by a metal mesh.

The heated zones of the shells can be more closely established by the proper use of cooling air jets. Cooling air jets are also directed upward under the shell skirt to control the temperature of the metal-glass button-stem seal.

The glass button-stem header with the mount cage on it is welded onto the shell flange before delivering the tube to the exhaust machine. However, on some metal types, a lead is brought out through the top of the shell. The seal for this lead is made on the sealex by means of a glass bead.

Although the shell is raised to a temperature of 1150 C, or higher, the working elements of the tube do not reach temperatures above 500 to 600 C. The shell-heating period is comparatively brief and uses only five exhaust positions; at 540 tubes per hour, it takes only 35 seconds; at 720 tubes per hour, it requires only 25 seconds. During this brief heating, the tube is evacuated in only three of these positions. Position No. 3 is used for the leak detector; position No. 4 is open to air; re-evacuation begins again on position No. 5. Under this schedule, each metal tube is pumped down from atmospheric pressure twice, instead of just once as is the case for all glass types.

This procedure is peculiar to metal tubes and makes it possible to admit air after the tube elements are heated. In the process, all traces of hydrocarbon or other contaminants are apparently burned up. This feature is essential when the temperature of tube elements during the exhaust process is kept so low.

Another feature is that the internal shell surface is oxidized. Some of this oxidized surface is reduced later in the exhaust cycle by gases released by the mount structure and cathode. The resultant matte sur-

face, having been degassed, then becomes a potential getter because of its finely divided or spongy nature. Its dull surface also readily absorbs heat reflected to it from the anode, heat which is transferred to the shell and then radiated to the surrounding air by the dull-black painted outer surface of the shell. As a result, the anode operates at a lower temperature than would be obtained with a bright internal shell surface.

All metal tubes use a batalum-type getter which is degassed and flashed by low ac voltage during the exhaust cycle. The getter is connected in series with the lead which grounds the shell to a base pin. The pellet getter placed in the shell dome is used in addition to the batalum getter only in high-dissipation types.

Electron bombardment of grids No. 1 and No. 2 is employed in such tubes as the 6F6 and the 6L6 to attain better degassing of these electrodes.

Sealex Exhaust of Gas Tubes

Some gas-filled types justify a special sealex setup which exhausts automatically and fills tubes to a closely controlled optimum gas pressure.

In some instances, the pressure may be as low as 100 microns, while in others it may be almost 100 mm Hg. These two extremes in pressure require somewhat different techniques of introducing the gas into the tube. Because the sealex-processed tube is hot when the gas is introduced, the gas has to be injected at a higher pressure than would be required on manual or trolley exhaust systems where the tube is cooled to room temperature before the gas is introduced.

If the gas is to be introduced at a relatively high pressure (10 millimeters or more), the gas is fed directly to position No. 14 on the machine. A T-connection is provided between the position and the gas supply so that the entire gas-feed system can be evacuated frequently.

The inlet gas pressure is controlled by two diaphragm-type regulators; one at the tank valve reduces the pressure to approximately 5 psi; the other, which has a sensitive diaphragm of much larger diameter, reduces the pressure from 5 psi to about 10 mm Hg, absolute.

However, if the tube is to be filled to the 100-micron level, a "gas-dosing valve" (originally designed by the author) must be connected into the line between the low-pressure regulator and the tube. This valve (used in addition to the two diaphragm-type regulators) works on the apportioning principle. Specifically, a small volume of the gas at a pressure within the control range of the second regulator is allowed to expand into the much larger volume represented by the tube and the connecting system. After each tube receives its charge of gas and starts to index into the tip-off position, the gas-dosing valve causes the entire gas-feed system to be partially pumped out prior to the gas-filling of the next tube. This operation is very important in the production of high-quality tubes because it prevents the accumulation of contaminants in a system in which the gas flow is only a small volume at a low pressure. It also protects against the possibility of spoilage of the gas supply by a slowly leaking tube in the gas-filling position. Variation of either the valve volume or the pressure, or both, permits accommodation of a broad range of tube volumes and of gas-filling pressures.

Effects of Exhaust Tubulation Size

The bore of the exhaust tubulation is the most important dimension so far as its conductance or effect on pumping speed is concerned; its length is a close second. However, when the tip-off operation is considered, wall-thickness becomes a most important factor.

The formula $S = r^3/L$, where S is the effective pumping speed in liters per second, r is the inside radius of the tubulation, and L is its length (both dimensions in millimeters), indicates that if the inside radius could be doubled, the effective pumping speed would be increased by a factor of eight; if the tubulation length could be halved, the pumping speed would be doubled.

Such radical changes, however, would cause considerable difficulty at tip-off. The large diameter tubulation tends to "suck in" and gives a tip that is very poor in quality, strength, and appearance. Increasing the wall-thickness would improve the situation and give a better-quality tip, but would increase the time required for the tipping operation.

A shorter tubulation is rather impractical at present, except possibly on subminiature-tube types.

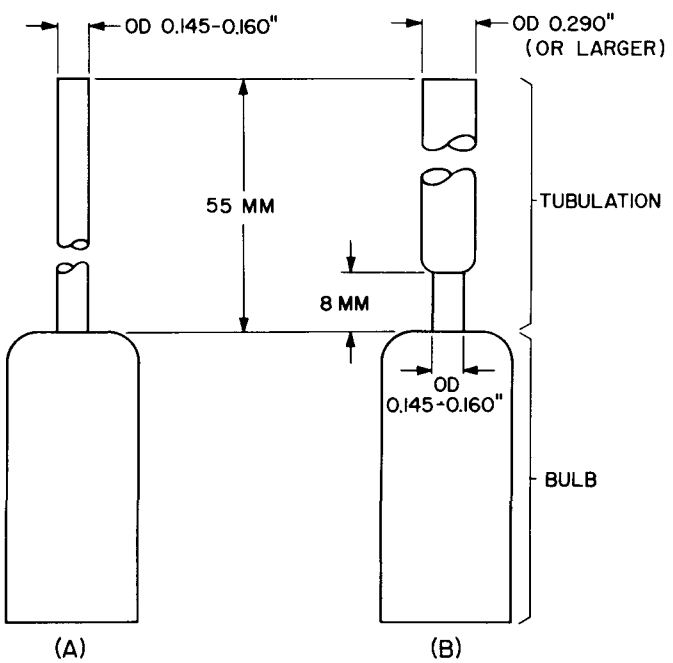


Figure 2. Bulb Tubulation: (A) regular miniature tubulation; (B) larger diameter constricted tubulation

A method for improving the exhaust tubulation situation uses a so-called constricted tubulation (Fig. 2) which is actually a tubulation having most of its length of expanded diameter. Such a tubulation could be between 1/4 and 3/8 inch in outside diameter for most of its 55-mm length. Only about 8 mm of its length, where it joins the bulb, would have the diameter and wall thickness of tubulation presently used for miniature tubes. This arrangement would assure the same tip quality now achieved with no new problems, but the pumping speed would be about seven times greater. This short small-diameter or constricted portion, and the long large-diameter section, along with its appropriate port rubber diameter (or better still, the use of a compression-type port) would improve the over-all tubulation-and-port performance by a factor of five.

Unfortunately, this improvement does not mean that the over-all pumping time of the tube can be reduced by a factor of five; the factor of five applies only to that portion of the vacuum system comprising the tubulation and the port rubber. Other portions of the sealex vacuum system leading up to the port also impose some limitations on pumping speed and prevent the full benefit of the improvement from being achieved.

Other Methods of Exhaust

Practically all receiving-type tubes are exhaust-processed either on sealex or trolley, as previously described. However, some other methods or variations have been used, or are in use, or are being investigated.

Sealex: Multiple Tubes per Port. With this method, two or more tubes are exhausted simultaneously on each port. This procedure has been used recently on certain special tube types. In general, the types involved are those that for particular reasons are processed at very slow machine speeds, e.g., index speeds of 120 tubes per hour, or less.

Obviously, a leak in any one of the tubes on the same port will affect the others on that port or, if the leak is large enough, will cause all of the tubes on that port to be shut off from the vacuum system.

The feature of this method is that it achieves a longer period of exhaust processing for a given number of tubes per hour output.

Batch Exhaust. The term batch exhaust, when applied to conventional tube types, refers to the simultaneous exhaust processing of a large number of tubes, from one hundred to perhaps one or two thousand.

Suppose that two thousand tubes ready for exhaust were loaded into a chamber which was then closed and evacuated for one hour. That would be batch exhaust at a rate of two thousand tubes per hour. Or the batch might be processed in a number of smaller chambers, either simultaneously or in rapid sequence, to exhaust the same total quantity per hour.

Basically, the object of such methods would be the processing of more tubes per hour with more time for each of the various processing steps. While such methods would attain these objectives, there are practical problems in that each tube in the batch would have to be given individual treatment. For example, the heaters must be lighted and the tubes must be closed by tip-off or other means. With conventional exhaust methods, the equipment required becomes very complex. In addition, separate heating coils, separate lighting contacts, and separate closure devices in each chamber would be costly. However, if the individual processing steps could be eliminated, then the batch method has interesting possibilities. These are being explored with the nuvistor design of tubes.

Nuvistor Exhaust Techniques. Nuvistors are a recent development comprising small metal-ceramic tubes. The design is such that the tubes require no heater lighting, no getter flashing, and no tip-off operation during the exhaust procedure.

These tubes are small and thimble-shaped and have ceramic-wafer stems at one end; they are degassed, their cathode coatings are broken down, and their stems are brazed to the shell (another method of closure or tip-off) by subjecting the tubes to heat in a high-vacuum chamber. Many different arrangements are being explored but the technique offers many benefits for the design and processing of future tubes.

REFERENCES

1. Metson, G. H., "Life of Oxide Cathodes in Modern Receiving Valves," Res. Rep. No. 12944, P. O. Engineering Dept., London, 1951
2. Private communication from Dr. George Sofer of National Research Corp.
3. Code I1
4. Code I2
5. Code I3
6. Jansen, C. G. J., and Venema, A., "A McLeod Manometer with Prescribed Volumes for Use as a Standard Instrument," Vacuum, Vol. 9, No. 3/4, pp. 219-229, July/Sept. 1959
7. Veis, S., "The Measurement of the Pressures of Various Gases by Means of a Pirani Gauge," Vacuum, Vol. 9, No. 3/4, pp. 186-189, July/Sept. 1959
8. Korner, R. L., and Rieger, G. N., "Thermistor Checks Tube Exhausting," Electronic Industries and Tele-Tech, Vol. 16, No. 1, January 1957
9. Todd, B. Johnson, "Outgassing of Glass," Jour. of Applied Physics, Vol. 26, No. 10, pp. 1238-1243, October 1956

BIBLIOGRAPHY

Articles

Metson, G. H., "The Physical Basis of the Residual Vacuum Characteristic of a Thermionic Valve," British Journal of Applied Physics, Vol. 2, No. 2, pp. 46-48, February 1951

Metson, G. H., "Vacuum Factor of the Oxide Cathode Valve," British Journal of Applied Physics, Vol. 1, No. 3, pp. 73-77, March 1950

Neumann, R., "High Vacuum Pumps; Their History and Development," Electronic Engineering (British), Jan., Feb., Mar., Apr., May 1948

Norton, F. J. and Marshall, A. L., "The Degassing of Metals," Metals Technology, Vol. 11, No. 1, T.P. 1643, 21 pp. Jan. 1944
della Porta, P., "Performance Characteristics of Barium Getters with Particular Reference to their Application in Thermionic Valves," Vacuum, Vol. IV, No. 3, pp. 284-302, July 1954 (Published Feb. 1957) (good reference list)
della Porta, P., "Performance Characteristics of Barium Getters at Elevated Working Temperatures of the Valves," Vacuum, Vol. IV, No. 4, pp. 464-475, October 1954 (Published June 1957)
Reid, R. B., "Water Vapor in Gas-Filled Lamps," Sylvania Technologist, Vol. IV, No. 4, October 1951
Stoll, S. J., "Electron Tube Gas," Army Ordnance Diamond Fuse Laboratory, T.R. 152, 1955
Stoll, S. J., "Origin and Analysis of Gas in Electron Tubes," British Journal Applied Physics, Vol. 7, pp. 94-96, 1956
Turnbull, J. C., "Electron Tube Materials, Gases and Getters," RCA Misc. Reports, Quarterly 1959
Wagener, S., "A Method for Measuring Efficiency of Getters at Low Pressures," British Journal of Applied Physics, Vol. I, No. 9, pp. 225-231, Sept. 1950
Wagener, S., "The Use of Getters for the Production of Very High Vacua," Vacuum, Vol, III, pp. 11-23, 1953
Wagener, S., "Adsorption Measurement at Very Low Pressures II," Journal Physical Chem., Vol. 61, p. 267, 1957
Weingarten, H. C. and Kennedy, S. W., "A Review of Modern Vacuum Pumps in Electronics Manufacturing," Advances in Electronics, Vol. 5, pp. 213-246, 1953, Academic Press Inc., (mech. and vapor pumps;sealex schematics; C.R. tube; power tube)
Industrial Electronics Reference Book, Westinghouse Elec. Corp., Chapt. 5, pp. 62-67, Wiley, New York, 1948, (oxide K, grid, anodes, envelopes, exhaust procedures pp. 65-66)

Periodicals

British Journal of Applied Physics (The Institute of Physics, London)
Journal of Applied Physics (Amer. Institute of Physics, N.Y.)
LeVide (French) (Societe Francaise des Ingenieurs et Techniciens du Vide, Paris)
Vacuum (British) (Pergamon Press, N.Y. and London)
Vakuum-Technik (German) (Rudolf A. Lang Verlag, Berlin)

Books

Braddick, H. J. J., The Physics of Experimental Method, 1954 (Wiley)
Benjamin, M., Modern Receiving Valves: Design and Manufacture, M. Benjamin, C. W. Cosgrove, and C. W. Warner, London, 1937, (pp. 401-439, reprinted from I.E.E. Journal, Vol. 80, No. 484, Aug. '37)
Dunoyer, L., Vacuum Practice, Bell & Sons, London, 1926
Dushman, S., Scientific Foundations of Vacuum Technique, Wiley, N.Y., 1949
Espe, W., Werkstoffe der Elektrotechnik, Akademie Verlag, Berlin, 1954
Guthrie, A. and Wakerling R. K., Vacuum Equipment and Techniques, McGraw-Hill, 1949
Hermann, G. and Wagener, S., The Oxide Coated Cathode, Chapman & Hall, London, 1951
Holland, L., Vacuum Deposition of Thin Films, Chapman & Hill, London, 1956
Jnanananda, Swami, High Vacua; Principles, Production and Measurement, Van Nostrand, N.Y., 1949
Kohl, Walter H., Materials and Techniques for Electron Tubes, Reinhold Pub. Corp., N.Y., 1960
Langmuir, I., Phenomena, Atoms and Molecules, Philosophical Library, N.Y., 1950
Leck, J. H., Pressure Measurement in Vacuum Systems, Inst. of Physics, London, 1957
Littman, M., Getterstoffe und ihre anwendung in der hochvakuumtecnik (gasaufzehrung durch metalldampf), (Leipzig, C. F. Winter), 1938
Newman, F. H., The Production and Measurement of Low Pressures, (E. Benn, London), 1925
Noelcke, C. L., Deterioration Mechanisms in Electron Tubes, Arinc Research Monograph No. 6, 1958 (Extensive listing of bibliographic material applicable to tube components, processing and operation)
Reimann, A. L., Vacuum Technique, Chapman & Hall, London, 1952
Strong, J., et al, Procedures in Experimental Physics, Prentice Hall, 1938
Vacuum Physics (a symposium of the Institute of Physics in Birmingham, England, June 1950), Journal of Scientific Instruments, Supplement No. 1, 1951
Vacuum Symposium Transactions, (1954, 1955, 1956, 1957, 1958, 1959), Pergamon Press, London
Van Atta, C. M., The Design of High Vacuum Systems, Kinney Mfg. Co., 1942
Yarwood, J., High Vacuum Technique, Chapman & Hall, London, 1955

Getters

C. H. Thomas

Lancaster

A thermionic vacuum tube requires that its pressure be reduced for most applications to a range of 1×10^{-6} to 1×10^{-7} millimeters of mercury (mm Hg) for stable operation and good life. Getters are used to assist the tube-exhaust equipment in obtaining the necessary low tube pressure. The first purpose of a getter is to clean up residual gas after exhaust and thus to permit a reduction in the time the tube must be connected to an exhaust system. Typical pressures in exhausted tubes prior to action by a getter are 1×10^{-1} mm Hg for a type 807, 1×10^{-4} mm Hg for a type 6146 and kinescope picture tubes, and 5×10^{-7} mm Hg for large power tubes.

The second purpose of a getter is the sorption of the gases which are evolved during the use of a tube through bombardment of its internal parts by electrons and the resultant heating of the parts and the tube envelope.

The use of getters has made it possible to mass-produce good tubes at low cost.

For clarification of text which is to follow, several terms are now defined. Getter flash is the term used to describe the deposition of getter material on the glass wall or other parts of a tube by heating a source of the chosen material to the desired temperature at which some of the material will vaporize rapidly. Small amounts of gas are measured in units of liter-microns. A liter micron is the amount of gas contained in a liter volume at a pressure of one micron (1×10^{-3} mm Hg) at 0 C.

GETTER ACTION

Getter action can be obtained in different ways. These include the following:

1. Chemical

   a. Flashed Metals. Active metals of the alkali and alkali-earth families in cold, flashed conditions will react chemically to some degree with the following gases usually found in tubes: oxygen, carbon monoxide, carbon dioxide, nitrogen, hydrogen and water vapor. Certain amounts of cesium vapor are gettered by alloying with tin or Aquadag.* Barium vapor is absorbed to some degree by alloying with gold or silver.

   b. Heated Metals. In recent years, the use of heated zirconium and titanium as getters has had much discussion. We know when these metals are heated from 700 to 800 C in oxygen that large amounts of oxygen are taken up. The reaction for oxygen is mainly non-reversible. These metals, in strip form, when heated from 600 to 700 C will also take up lesser amounts of carbon monoxide, carbon dioxide, and nitrogen. This sorption is chiefly non-reversible.

2. Adsorption. All surfaces will adsorb gases according to the area of their exposure, the degree of their degassed state, and their temperature. Degassed graphite will adsorb certain gases to some degree, but coconut charcoal, cooled to liquid-air temperature, will adsorb large amounts of all gases. Degassed copper foil at liquid-air temperature will adsorb certain gases to some degree.

3. Ion Bombardment. All gases, including rare gases, are driven into metal by ion bombardment provided the energy of the ion is high enough. The lighter gases are cleaned up more easily. This process occurs in all vacuum and gas filled tubes. When a tube is well exhausted, the pressure in such a tube may be decreased during life by clean up of the gas through positive ion bombardment of the metal parts and glass walls.

4. Sputtered Metal. All gases will be absorbed to some degree by metal particles sputtered from a cold metallic cathode under the influence of a high electric field. The rates of clean up of common gases are increased when active metals such as titanium and zirconium are used as cathodes.

HISTORICAL TYPES OF GETTERS

Information on getter development can probably best be presented by giving the getters more or less in their chronological order of use with a short description of each.

Phosphorus

Phosphorus was one of the earliest getters used in incandescent lamps to absorb water vapor. The tungsten filament was dipped into a solution of red oxide of phosphorus before mounting.

*Registered Trade Mark of Acheson Colloids Company

Magnesium

Magnesium was used in the form of ribbon which was attached either to the plate itself or, for high-frequency heating, to a metal flag. The ribbon was flashed after the exhaust process was completed.

Barium Azide

One of the first methods of producing barium as getter was the use of barium azide, BaN_6, which was applied in a water solution to a part of the tube which could be heated by high frequency. This compound was partially decomposed during tube bake-out with evolution of nitrogen. Complete decomposition did not occur until high-temperature heating with high frequency power when nitrogen was evolved as the barium was vaporized. The explosive and poisonous potential of this compound limited its use.

Misch Metal

Misch metal is an alloy of the rare-earth metals, principally cerium and iron, or cerium, iron, and magnesium. This alloy, which is relatively stable in air, will flash when heated in vacuum at about 1050 C. The use of Misch metal as a getter was very limited. However, it is presently used in certain voltage regulators where the Misch metal is sputtered to the cathode surface to lower the voltage drop. Although some gettering of deleterious gases occurs during this sputtering, an auxiliary barium getter is used in these voltage regulator type tubes.

Barium

An early source of barium was a wire made with a core of barium, and clad in iron, nickel, or copper. The wire most often used was an iron sheath which was ground thin on one side as shown in Fig. 1. The clad material was cut into sections with closed ends to minimize exposure of the barium. These sections were attached to metal flags by placing the section between the flag and nickel mesh which was spotwelded to the flag surface. The flag could then be heated to 1000 C on exhaust to give a flash of barium. The cost of this getter and its deterioration on prolonged exposure to moist air prevented its extensive use. However, the iron-clad barium can still be purchased in lengths of 8 to 12 inches. These lengths can be used in the form of a coil which can be flashed by passage of sufficient current through the wire to heat it to 1000 C. This form of getter is used in RCA Super Power tubes to give several hundred milligrams of barium flash.

Barium-Aluminum-Magnesium Pellet

The first getter to have extensive use was that made from a barium-aluminum-magnesium alloy of approximately 43 per cent barium, 20 per cent magnesium, and 37 per cent aluminum. This alloy was ground into powder which was pressed into a small pellet. A pellet was fastened to a metal flag which could be heated by high frequency during exhaust to a temperature of 700 to 1000 C where the magnesium and some barium would be flashed. The relatively poor ability of the getter flash area to absorb deleterious gases, and the difficulty of confining the getter flash area limited its use in receiving tubes.

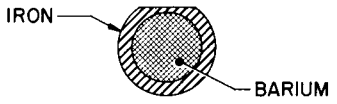


Figure 1. Cross-Sectional View of Early Barium Getter

Barium-Aluminum Pellet

A barium-aluminum alloy of approximately 50 per cent barium and 50 per cent aluminum was used to some degree in the form of a pellet made from powder. However, the difficulty of obtaining a uniform flashing temperature of 1050 to 1100 C from a metal flag restricted its use.

Batalum Getter[1]

The advent of the metal receiving tube required development of a getter that could be flashed with a low electric current through a ribbon or wire that could be attached to the leads of the tube. The barium was obtained by a chemical reaction of the barium oxide with tantalum at elevated temperature. In practice, the process is most conveniently carried out by coating a tantalum wire with barium and strontium carbonates, converting the carbonates to oxides at about 1000 to 1100 C, and then raising the temperature to about 1300 C at which temperature the barium and strontium metal evaporates. Barium and strontium carbonates were used because barium carbonate alone melts and would cause sputtering.

The disadvantages of this getter were the necessity of decomposition of the carbonates during exhaust and the high cost.

The getter was improved by making a barium berylliate compound, $BaBeO_2$, which could be applied to a tantalum channel or ribbon.

In the final form the barium berylliate powder was mixed with titanium powder. This material, applied in powder form to a molybdenum ribbon or channel, gives barium when the ribbon is heated to 1200 or 1500 C. The advantages of this getter are its wide range of flash temperature and its confined flash region; the disadvantages are the presence of metallic particles in the tubes and the relative inability of the flash to give consistent oxygen absorption.

In spite of these disadvantages, the Batalum getter had very wide use for a number of years and is still used in metal tubes.

Kic Getter*

The "Kic" getter, which appeared in the early nineteen forties, is a barium-aluminum alloy containing

*Registered Trademark of Kemet Company

approximately 50 per cent of each metal. The alloy, which is stable in air, is used as a wire which is clad with nickel or stainless steel. The wire is ground on one side to expose the alloy for ease of flashing (see Fig. 2). The wire is welded in single or parallel strands to a metallic loop for heating by a high-frequency source. When a double-strand getter loop is heated by high frequency the outer channel will become hotter than the inner because of the higher induced current resulting from the larger enclosed loop.

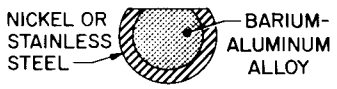


Figure 2. Cross-Sectional View of Kic Getter

Barium of high purity can be obtained in a wide range of amounts from this type of getter which is available in different lengths and wire diameter. Although this type of getter has very little occluded gas, some hydrogen is given off during the flashing period.

King Getter*

The King getter consists of pellets of an alloy of 50 per cent barium and 50 per cent aluminum imbedded in a strip of nickel or stainless steel. A length of this strip, with the desired number of pellets, is welded to the legs of a metal loop for high-frequency flashing. The quality of the getter flash is similar to that of the "Kic" type, but less hydrogen is evolved during flashing.

Ring-Type Getter

Within the past few years the ring-type getter has come into increased use primarily because of its reduced cost. A ring of metal of the desired dimensions has the desired amount of getter powder pressed into the channel which usually has a cross-sectional design of the form shown in Fig. 3. The getter powder may be of two general classes: exothermic and non-exothermic reacting. The composition of the exothermic powder varies somewhat with the different manufacturers. One analysis shows approximately 35 per cent barium, 30 per cent aluminum, and 35 per cent nickel. This mixture reacts exothermically when it is heated to about 1000 C to give about half its possible barium yield. A longer heating time at a higher temperature is necessary to obtain higher percentage barium yield. The exothermic reaction of this getter makes the control of the flash difficult when full barium yield is required.

The ring-type getter is also used with non-exothermic getter powder which is made from a barium-aluminum alloy of approximately 50 per cent barium and 50 per cent aluminum. This powder is used in ring-type getters primarily for tubes which require larger amounts of flashed barium.

*Registered Trademark of King Company

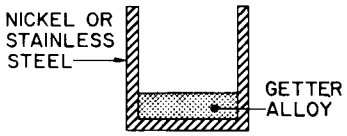


Figure 3. Cross-Sectional View of Ring-Type Getter

Zirconium and Titanium

Both zirconium and titanium, in flashed condition, will take up gases. Moreover, both metals are very active during the vaporized state. Both metals are used in finely divided form or in strip form which can be heated during tube operation. The powder of the metal or its hydride is usually applied to the outside of the plate of the tube where it can be well degassed during exhaust. During operation, the plates operate in the temperature range of 400 to 700 C; within this temperature range the powder will take up some of the common gases. Either metal is used in some tubes for gettering action where the metal itself can be kept at a temperature above 500 C.

There is very poor agreement in results of the various investigators on the efficiency of zirconium and titanium as getters for the common tube gases. It is known that zirconium or titanium powder, or hydrides of these powders, can be used as hydrogen generators with a reversible reaction over the range of about 350 to 700 C.. However, severe oxidation of the powder upsets this behavior with hydrogen.

The Vac Ion getter pump* uses titanium strip for electrode material which, at room temperature, getters the gases through ion bombardment and through chemical reaction with the titanium which is sputtered from the electrodes by high voltage ion bombardment.

Tantalum

Tantalum metal is used in certain high-power gas-filled thyratrons for anodes which operate at 500 to 800 C through electron bombardment. Under these conditions, the tantalum metal has some getter action for the usual tube gases — oxygen, nitrogen, carbon monoxide, and carbon dioxide.

Cer-Alloy**

Within the past several years, New Process Metals Company has supplied a new type of getter powder, Cer-Alloy, which is a mixture of cerium and thorium. Although laboratory tests show that this powder is effective for gettering of the ordinary gases when it is heated to the 500 to 600 C range, its susceptibility to damage from moisture and high-temperature exposure during the sealing operation has greatly limited its use in vacuum tubes.

*Registered Trademark of Varian Associates

**Registered Trademark of the New Process Metals Company

Cerium-Thorium Alloy

Phillips Company of Holland has made an alloy of cerium and thorium which can be made into wire and strip which allegedly acts as a bulk getter to take up gases near room temperature.

Although extensive claims have been made for the gettering action of this material, the manufacturer has not released material for test.

EFFICIENCY OF GETTERS

One of the most extensive early works on measurement of getter efficiency for absorption of common gases was that of Ehrke and Slack[2]. Their method of measurement involves the introduction of the gas to be tested into a getter-flashed bulb which can be isolated from the pumps by a suitable cutoff. The gas is introduced at a pressure in the micron range and then the change of pressure with time is noted. The reference article gives a tabulation of the results for common getters for absorption of gases with both bright and diffuse flash surfaces. A diffuse surface was produced by flashing each getter in argon at a pressure of 1 to 3 mm Hg. (At RCA, argon at a pressure of 10 mm Hg has been found to be more efficient.) The authors report much higher efficiency of gettering for the diffuse flash. This effect is what one would expect because of the higher porosity of the flash produced by flashing the getter in argon.

Recent extensive studies of gettering at low pressures have been done by Wagner[3,4], P della Porta[5], and Morrison[6]. These investigators use two ion gauges connected by a suitable capillary to measure the flow of gas by the use of Knudsen's Law for the flow of gases at low pressure. This formula is:

$$G = F\frac{P_m - P_T}{P_T}$$

where G is the rate of absorption of gas by the getter in liters per second as measured by the flow of gas from the manifold through the capillary; F is the calculated speed of gas flow through the capillary; P_m is pressure, in mm Hg, in the manifold; and P_T is pressure in, mm Hg in the tube with the getter flash. These investigations have been carried in the pressure range of 10^{-5} to 10^{-9} mm Hg with special consideration for correction for the absorption of gas by the glass walls and by the operation of the ion gauges.

These investigators have used a getter-flash area of 5 to 10 square centimeters with a barium deposit of 0.05 to 0.10 milligrams per square centimeter. Wagener[3] has reported an initial speed of clean-up for oxygen of 11,000 cubic centimeters per square centimeter per second.

Wagener continued his test only until the speed of absorption was decreased to 5 to 10 cubic centimeters per second. He reports the time required (of the order of minutes) as the speed of gettering for an oxide coated cathode. He claims that the speed of gettering during the first few minutes can be used to predict the behavior of the getter during tube life.

Usual RCA factory practice is to flash a getter to give a deposit of 0.05 to 5.0 milligrams per square centimeter. During life, the total area of the getter flash decreases as the barium is used up. We have also measured a loss of 50 milligrams of barium in the getter deposit in a color kinescope during 3000 hours life. In the light of this information, it is difficult to see how Wagener's data on initial speed of getter action can be applied to our factory problem. However, the past work on getters does show good agreement in three general areas:

1. The efficiency of getter increases with the effective surface area of the getter-flash deposit.

2. The efficiency of the getter is increased materially by flashing the getter in argon at a pressure of a few millimeters to give greater porosity.

3. The temperature of the barium getter deposit must be increased above 150 C to increase its adsorption efficiency significantly.

An extensive bibliography on gettering has been assembled by Bartin and Tocci[7].

During the past five years, fairly extensive tests have been made at Lancaster to determine the efficiency of getters for absorption of gases at relatively high pressure.

A schematic drawing of the all-glass system used for such measurements is shown in Fig. 4. The gases to be tested are bought in liter bottles. Each bottle G is separately connected to the system through two vacuum-grade stopcocks S_1 and S_2 spaced to have a volume between them of about 5 cubic centimeters for dosing the system. One or more of the gas bottles can be connected to the system reservoir PV which has a volume of several liters and which includes McCleod gauge MC before the first mercury cutoff C_1. This cutoff connects the gas reservoir to the tube part of the system which includes a cold trap CT and the tube T for the getter. The tube can be baked out by means of an electric furnace. A second mercury cutoff C_2 connects the entire

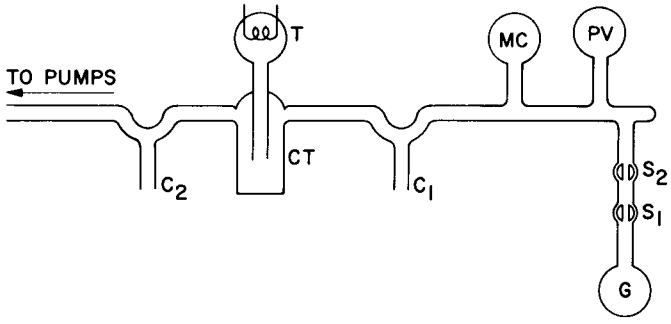


Figure 4. Apparatus for determining Getter Efficiency (G-gas bottle; S_1, S_2—stopcocks; PV-system reservoir; MC—McCleod gauge; C_1, C_2—cutoffs; CT—cold trap; T—tube containing test getter)

system to a mercury-diffusion pump and a mechanical pump.

When a getter is to be tested, it is sealed into a suitable tube T, which is then sealed to the system. The entire system is pumped through C_1 and C_2. The tube T is first baked at a temperature of 350 to 400 C and then is allowed to cool to room temperature. Cutoff C_1 is closed after which the reservoir system is filled to the desired pressure (100 to 3000 microns) from one of the connected bottles by means of the dosing stopcocks S_1 and S_2. Cutoff C_1 can be adjusted to give the desired pressure so that the desired amount of gas is in the reservoir. Liquid-air or dry-ice refrigerant must be applied to the cold trap to freeze out mercury vapor which affects the gettering properties of barium.

The sequence of operations is as follows: the getter in T is degassed and flashed; the getter flash is allowed to cool to the desired temperature; cutoff C_2 is closed; cutoff C_1 is opened to allow the gas in PV to expand into new volume V_1; the new pressure P_1 is calculated for the condition before getter flash; pressures of P_1 are measured on the McCleod gauge at intervals of 1, 2, 3, and 5 minutes after getter flash, and as much longer as is desired. Measurements may even be continued up to 24 hours. Usually, the getter is practically used up or is inactive after one hour for a barium deposit of a fraction of a milligram per square centimeter. However, the clean-up may extend for days with oxygen gas and a barium deposit of 15 to 20 milligrams per square centimeter.

Typical results are given in Fig. 5 for the clean-up by barium of oxygen and of a mixture of oxygen and carbon monoxide. Twenty-five milligrams of barium flashed over 100 square centimeters cleaned up 1900 liter microns of oxygen. This is equal to 76 liter microns of oxygen per milligram of barium per hour. If it is assumed that the clean-up reaction is chemical, then $2\,Ba + O_2 \rightarrow 2BaO$. The theoretical value is 68 liter microns of oxygen per milligram at 26 C. The author has found a range of 30 to 110 liter microns per milligram of barium for oxygen. No proven explanation for these values is available, but their validity has been checked many times. The curve for absorption of the

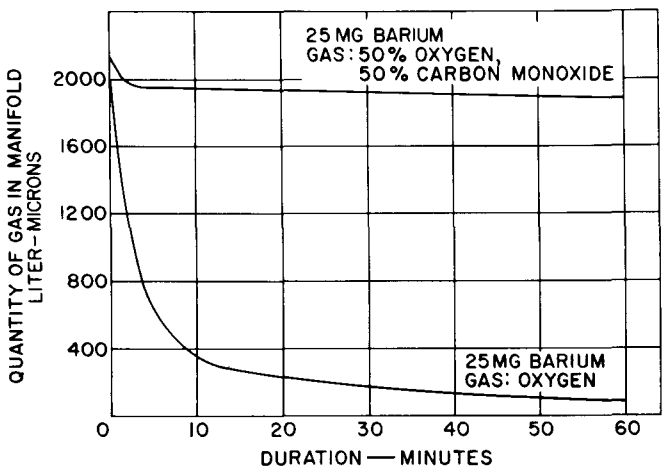


Figure 5. Gas Absorption by Barium Getter

mixture of oxygen and carbon monoxide shows that 25 milligrams of barium flashed over 100 square centimeters cleaned up 14 liter microns of gas per milligram of barium in less than an hour. The reaction was practically complete in one hour. Further exposure of several days failed to provide any significant additional clean-up. However, when the carbon monoxide content was reduced to 10 per cent, the clean-up continued in a way similar to that for oxygen. The results for several mixtures of oxygen and carbon monoxide, and for a mixture of oxygen, carbon monoxide, and hydrogen, are given in Table I.

Table I

Effectiveness of Getters

Getter	Oxygen per cent	Carbon Monoxide per cent	Hydrogen per cent	Absorption liter microns per milligram
Barium	100	- -	- -	30-110
Barium	90	10	- -	40- 50
Barium	75	25	- -	10- 20
Barium	50	50	- -	6- 15
Barium	0	100	- -	8- 15
Barium	25	25	50	22
Cesium	50	50	- -	22
Potassium	50	50	- -	25
Titanium (flashed)	50	50	- -	15- 17

When carbon monoxide is admitted over a getter flash of barium, the clean-up of the gas is rapid for a few minutes after which the rate is very slow. This decrease is thought to be due to the formation of a protective layer on the surface of the getter flash. Mixtures of oxygen and carbon monoxide, with the carbon monoxide content 25 per cent or greater, react in the same manner as shown in the top curve of Fig. 5.

An experiment was carried out in which a gas mixture 50 per cent oxygen and 50 per cent carbon monoxide was exposed to a barium getter flash for about 20 minutes at which time the rate of clean-up of the gas was very low. This gas mixture was pumped out. When oxygen was readmitted over this getter flash the rate of clean-up of the oxygen was rapid. This indicates that a certain partial pressure of carbon monoxide gas must be present to maintain the barrier layer on the getter surface.

Fig. 6, which gives the curve for the clean-up of oxygen by a very heavy flash of barium, shows a clean-up of 81 liter microns per milligram of barium after 100 minutes, and 108 liter microns at 72 hours. The intervening behavior is unknown as indicated by the broken line.

Fig. 7 shows the clean-up of a mixture of 50 per cent oxygen and 50 per cent carbon monoxide by titanium. The top curve shows the clean-up by 20 milligrams of titanium flashed over an area of 200 square centimeters. At 30 minutes, the clean-up is 15 liter microns per

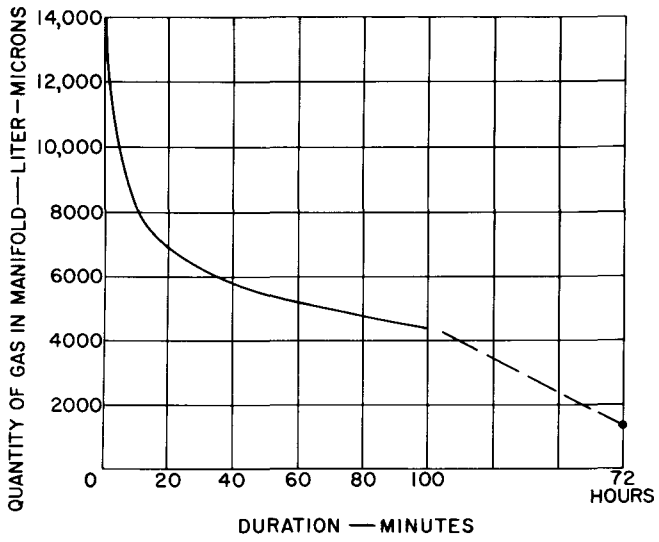


Figure 6. Gas Absorption by 117-Milligram Barium Getter

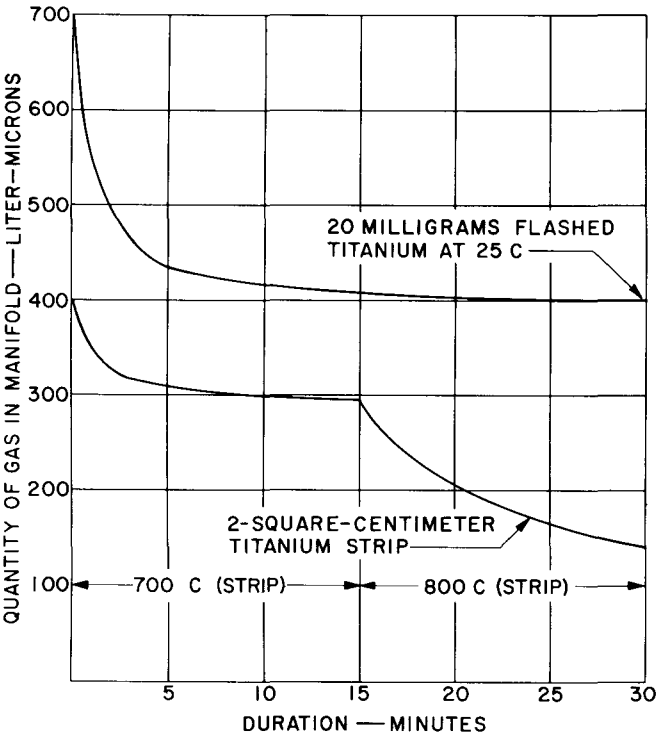


Figure 7. Absorption of a Mixture of 50 Per Cent Oxygen and 50 Per Cent Carbon Monixide by Titanium

milligram of barium. The bottom curve shows the behavior of a 2-square-centimeter strip of titanium heated in the presence of a mixture of oxygen and carbon monoxide. At about 700 C the strip oxidized badly but cleaned up gas fairly well for a few minutes; when the temperature of the strip was increased to 800 C the clean-up increased rapidly as shown.

Although many tests of heated strips of zirconium and titanium have been made in the various gases, the results have been extremely inconsistent. As a result, no reliable data are available.

Table I summarizes some RCA getter data. The ranges of clean-up efficiency are due to differences in area of flash deposit, in the type of getter, and in the thickness of the flash deposit.

EFFICIENCY OF GETTER FLASH

Preflash

In the exhaust of most receiving tubes, economy requires that the getter be flashed in a few seconds on the exhaust machine before tipoff. This practice gives satisfactory tubes. In the exhaust of low-pressure gas tubes on a sealex machine equipped for dosing, a getter must be flashed before dosing the tube to remove residual gases so that gas at the correct dosing pressure may be put into the tube.

Postflash

A better vacuum is obtained in a vacuum tube by postflash of getter after tipoff because the vapor stage of the getter is most effective for gettering residual gases.

Postflash of getter is used in medium-priced power tubes and in medium-pressure gas-type voltage-regulator tubes where the filling gas can be purified by postflash of the getter.

Because most getters have residual gases, they should be degassed on exhaust, if possible, by heating the getter to 800 or 900 C. With Kic getters, the residual gas after getter flash is mostly hydrogen which is not readily taken up by the barium getters. However, the hydrogen is rapidly cleaned up during operation of the tube by ion bombardment produced when the anode current ionizes the hydrogen.

Flashing in Argon

A black, porous getter surface is produced by flashing the getter in argon at a pressure of 5 to 10 millimeters. This surface is many times more reactive than the bright surface of a flash deposit produced in high vacuum. This procedure is used in Lancaster for the glass-enclosed getters which are placed in the large super-power tubes.

DETERMINATION OF BARIUM IN GETTER FLASH

The simplest method for determining the amount of barium in a getter flash deposit is to titrate a water solution of the getter-flash residue with hydrochloric acid using a methyl purple indicator. Extreme care must be taken to dissolve all the residue of the getter from the glass walls. Using some of the titrating acid and a small rubber spatula for removing the residue makes for more accurate results.

The amount of barium can also be determined by the flame spectrophotometer method perfected by Bertin[8]. This method is especially useful where the getter flash is in the range of 1 to 3 milligrams of barium, as it is in small receiving tubes.

REFERENCES

1. Lederer, E. A., and D. H. Wamsley, "Batalum," RCA Review, Vol. 2, pages 117-123, 1937

2. Ehrke, L. F., and C. M. Slack, Journal Applied Physics, Vol. 11, pages 129-137, 1940

3. Wagener, S., Journal of Physical Chemistry, Vol. 60, pages 567-576, 1956

4. Wagener, S., Journal of Physical Chemistry, Vol. 61, pages 267-271, 1957

5. della Porta, P., Vacuum, Vol. IV, No. 4, pages 464-475, 1954

6. Morrison and Zetterstrom, Journal of Applied Physics, Vol. 26, pages 437-442, April, 1955

7. Code G1

8. Code G2

Aging and Stabilization

G. Wolfe

Harrison

Aging is the process of treating tubes which have oxide-coated cathodes, after exhaust, to activate the coating for maximum thermionic efficiency and to clean up residual gas. Stabilization refers to the process of establishing conditions within the tube for long and uniform emission performance and, therefore, is inherent in all aging processes. The term, however, as usually used, is applied to special processing of longer duration than is employed for conventional aging.

AGING

BASIC CONSIDERATIONS

The exhaust process and the aging process are interdependent to an important degree. The kind of exhaust equipment used and its rate of operation determine certain requirements for the aging process. For example, vacuum-type receiving tubes are usually exhausted on sealex (combination sealing and exhaust) machines at speeds ranging from 200 to 1200 tubes per hour. For a 16-head sealex machine operating at a commonly used speed of 600 tubes per hour, the total time a tube is connected to working heads is only about one minute. It is during this short time that tube parts must be brought up to temperature, the tube evacuated, and the cathode "broken down" (nitrocellulose driven off and carbon dioxide and barium oxide set free).

The process of activating the indirectly heated oxide-coated cathode, although not completely understood, involves the eventual production of donor states in the material originally applied to the cathode sleeve as barium carbonate. (More accurately, the cathode coating is a co-precipitated mixture of barium, strontium, and calcium carbonates, or occasionally barium and strontium carbonates, but only barium will be discussed.) The cathode sleeve is essentially nickel plus certain trace-element alloying agents which are added to enhance the workability of the material and the mechanical stability of the sleeve, and to permit chemical "reduction" of the cathode coating. The coating, as applied, consists of barium carbonate in a nitrocellulose binder together with vehicles to permit handling and spraying. The coating on the cathode that is eventually presented to the exhaust system is the barium carbonate and nitrocellulose.

When the cathode is heated ("lighted") during exhaust for approximately one minute, the nitrocellulose binder is driven off (at approximately 200 C) and the barium carbonate reacts under the applied temperature (1000 to 1250 C) to produce carbon dioxide and barium oxide. The carbon dioxide is pumped out, and the cathode coating is then only barium oxide, a material that will support thermionic emission, but only for a short period of time and with considerable instability.

Because it is necessary to produce donors in the oxide coating, additional processing is required; this processing consists mainly of thermochemical and/or electrochemical reactions. The thermochemical reactions are usually begun and partially completed during exhaust, but the major portion of this reaction must usually be performed independently from exhaust during the process known as "aging."

The reducing agents necessary for the production of donors during aging are contained in the cathode-sleeve alloy. The reducing agents include magnesium, silicon, and titanium. Other materials may also be used. The amount included of each material depends upon the eventual use of the tube type; there are "active," "normal, and "passive" cathode materials which are used in tubes having varying life requirements. This description is oversimplified, but will suffice for this discussion. Except when extremely long life is required (in which case a cathode material relatively close to pure nickel would be used), the cathode material can be assumed to include a reducing agent that can be called upon to aid in the production of donors during aging.

Because the chemical reduction depends upon the application of heat, the cathode could be activated by simply connecting the heater of the tube to a source of voltage and heating the cathode. However, although this method would produce donors and a high level of emission, and is in effect the first step in the conventional aging schedule, it rarely results in a stable tube. Additional processing is almost always needed to remove residual gases from other metal parts which, during exhaust, may not have been heated enough to completely degas them. Although rf heating during exhaust does, to a large extent, degas tube parts in which eddy-current effects occur efficiently, this degassing, because it occurs under varying vacuum conditions and is completed before final pressure is achieved at getter flash, is incomplete. Some parts, such as certain grids, may even be coupled poorly to the rf coil deliberately to prevent excessive heating and resultant deformation

and/or sublimation.

In any event, it is necessary to re-heat such parts through electron bombardment to liberate gases that would otherwise evolve during tube life as a result of the special surface effects from electron bombardment. Because the gases set free in this manner tend to cause transient cathode poisoning, this "clean-up" must be performed under accelerated and controlled conditions during tube manufacture to produce a stable circuit element.

THE AGING PROCESS

The Hotshot Step

The first step in the aging procedure is usually a thermal treatment known as the "hotshot" step. This step reduces the barium oxide to activate the cathode coating, and also provides a source of electrons to be used for element bombardment in later stages of the aging process. The time and temperature of the thermal treatment is limited by the heater, the cathode sleeve material, the cathode coating, the temperature and time required for optimum activation, and a time interval consistent with manufacturing feasibility. The usual practice is to apply a heater voltage of 150 to 200 per cent of rated voltage for 2 to 5 minutes. Low-temperature processing has many advantages and, in fact, may be necessary for highly active cathode materials which normally contain reducing agents that sublime readily. This sublimation can produce interelectrode leakage. Passive cathode materials can be subjected to higher processing temperature. For pure nickel cathodes, the only temperature limitations with respect to sublimation are those imposed by nickel alone. Other effects, such as migration of cathode coating, bowing of cathodes, and mica breakdown may limit maximum cathode temperature during the exhaust if they occur before the temperature limitation imposed by sublimation is reached.

The reduction of the cathode coating by the base metal produces donors first at the interface between sleeve and coating. During the continued activation, these donors diffuse to the surface of the coating; this diffusion is the goal of the reduction process. In the course of further activation by thermal reduction, however, the donors may be vaporized from the surface. Such evaporation has two bad effects: first, the rate of evaporation may exceed the reduction and diffusion rates and thus de-activate the cathode; second, the evaporated materials may sublime upon adjacent tube parts and may cause interlectrode leakage or contribute to undesirable thermionic emission from grids. An optimum aging condition is one which produces donors at a rate which permits complete diffusion into the coating but does not cause de-activation and evaporation of sublimates onto other tube parts. The importance of de-activation is lessened by the fact that there is a time-temperature range within which activation by reduction and diffusion may overcome de-activation by evaporation. Deposition of sublimates on adjacent tube parts, on the other hand, is usually not counteracted, and its occurrence is generally considered to be harmful.

Overheating the cathode at exhaust or aging may fuse the coating material. Short mount structures, where cathode-temperature variations may be extreme as a result of appreciable cooling effects at the ends of the cathode, are more susceptible to overheating. End-cooling effects may also be exaggerated by a tight fit of the cathode in the supporting mica. Processing schedules guided by the temperature of the ends of the cathode may lead to fusion at the center. Fusion of the coating makes activation extremely difficult in the fused area, and is usually reflected in slumping emission and poor emission at lower heater voltage.

Cathode-coating evaporation and fusion during exhaust and aging are related to the cathode coating. A fluffy coating is more prone to evaporation than a dense coating. However, a dense coating is more subject to fusion, particularly in constructions which introduce critical process-temperature problems.

During exhaust, the cathode temperature is raised to 1000 or 1250 C to break down the cathode coating, and gases are liberated from the heater, heater coating, cathode sleeve, and cathode coating. These gases are removed from the tube during the exhaust cycle. In a properly exhausted tube, the cathode coating is completely broken down and the gases liberated are completely removed. During aging of such a tube, however, the cathode might be heated to a temperature appreciably higher than that reached during exhaust. This higher temperature tends to liberate gases from the heater and cathode which had not been liberated at the lower temperatures on exhaust. Because these gases cannot be pumped out of the tube, they may either be adsorbed by the getter or combine with the donors being produced. Adsorption by the getter is undesirable in that the getter may be prematurely depleted; combining with the donors might cause temporary or permanent poisoning of the cathode coating, and render activation either extremely difficult or impossible. This consideration, then, imposes a limitation to maximum hotshot temperature beyond those implied by materials, construction, and application details.

The Grid Step

After a supply of donors is produced through initial thermal reduction, the next step in the aging procedure is stabilization of the emission and the degassing of tube elements under electron bombardment. In many tubes this is begun by a "grid step." The heater voltage is reduced from the hotshot voltage to a value about 20 to 40 per cent higher than rated voltage, and voltage is also applied to grids and plates. Thermionic emission current then flows to those elements. The grid and plate voltages are adjusted to provide dissipation in each electrode which is the same as or slightly higher than the dissipation which occurs during normal tube operation. Because it is the function of tube aging to remove gases which would be liberated at normal operating temperature by electron bombardment, the "same-as-or-slightly-higher-than" dissipation condition is used to assure complete normal-temperature degassing in a reasonable time in the production process.

The electrode voltages used during aging depend upon

the tube construction. For diode aging, positive voltage is simply applied to the plate until the appropriate plate dissipation is reached. In a triode, aging voltages are applied to both the grid and the plate. A triode grid normally operates at a negative potential and is, therefore, not heated appreciably, during tube use, by electrical conduction. As a result of its closeness to the cathode, however, the grid is heated by radiation from the cathode. Consequently, it must be degassed during aging to reduce the possibility of gas evolution during normal operation. Grid aging also cleans the grid surface of cathode-coating material deposited during breakdown and activation, and thus prevents grid emission. A low value of ac or positive dc grid voltage is applied to give a low level of dissipation. Because the application of positive grid voltage increases electron flow to the plate, grid and plate voltages must be adjusted interdependently to provide proper dissipation levels.

The conditions described for the triode also apply to the control grid and plate of a pentode or beam power tube. However, these tubes also contain a screen grid and suppressor grid (or beam-confining electrode). The screen grid normally conducts and dissipates heat in the same manner as the plate. The aging consideration, therefore, is the same; voltage is applied to the screen grid to provide dissipation that is the same as or slightly higher than the rated dissipation. The suppressor grid, although it is probably not degassed appreciably during rf heating at exhaust, does not conduct during tube operation and does not normally require aging. The same operating condition generally applies to beam-confining electrodes, although they are normally degassed to a greater degree during exhaust than is the wound suppressor grid. The suppressor grid is usually tied to the cathode during aging and suppresses secondary emission from the plate in the usual manner.

The second aging step, during which voltage is applied to grids and plates to degas them through bombardment, also stabilizes cathode emission. Because the cathode temperature is maintained at about 20 to 40 per cent above rated temperature, it receives a continuing thermal treatment which tends to counteract the poisoning effect of gases liberated from other electrodes. These gases combine readily with the donors produced in the previous step, and the continuing thermal reduction provides for re-activation of the cathode. This thermal reduction at a condition of electrical conduction probably enhances the production of donors through electrolysis. The process, however, is not yet fully understood. The stabilization of emission which occurs during this aging step is believed to be related to the high cathode current and to the resultant electrolytic dissociation of the coating ($2BaO \rightarrow 2Ba+O_2$).

The use of positive grid voltage is usually the most significant factor in determining tube conditions and limitations during the grid step. Although grid dissipation is maintained at a low level, there are considerations which limit the time in which such dissipation can be tolerated. The grid is of fragile construction and is often plated or clad with a material which sublimes readily. These plating or cladding materials are usually gold, silver, or similar metals which suppress thermionic emission from the grid. The time and conditions for the grid step, then, must be limited to prevent excessive mechanical deformation of the grid or excessive sublimation which could cause cathode poisoning. The possibility of over-activating (or de-activating) the cathode must also be considered. The high current may affect the cathode chemically. The electrolytic effect may produce oxygen, and its rapid or excessive evolution within the cathode coating may cause de-activation.

The Anode (or Plate) Step

The plate of the tube during the grid step operates at high current and low voltage because of the accelerating effect of grid voltage. The resultant heating of the plate approaches the effect of normal electron bombardment more closely than that produced by the rf heating stage. However, the normal electron bombardment at the plate surface during life takes place at high voltage and low current. Plate surface treatment under relatively high-velocity electron bombardment, at approximately rated dissipation, is necessary to effect surface degassing in the manner most nearly approaching normal operating bombardment characteristics. The next aging step, the "plate step," provides for most effective plate degassing.

In a triode or pentode, the control grid may be negatively biased, grounded, or even "floated" (least controllable) during the plate step. In a pentode, the screen grid must be considered in the same light as the plate. Because there is conduction to the screen grid, similar high-velocity bombardment at approximately rated dissipation is necessary for appropriate surface degassing. The suppressor grid, or the beam-confining electrode, is tied to the cathode during the plate step for the reasons mentioned during the discussion of the grid step.

The heater voltage may either be maintained at 20 to 40 per cent above the rated value during the plate step, or it may be reduced slightly from the value used during the grid step. Its only function at this stage is to keep the cathode at a temperature slightly above normal to achieve plate (and screen-grid in a pentode) dissipation at normal surface bombardment conditions. The cathode should be operated at a slightly elevated temperature to prevent cathode poisoning by gases liberated from the plate and screen grid. Transient poisoning of the cathode by gases liberated from electrodes during aging must be overcome by re-activation through continued or repeated thermal reduction and conduction in the cathode during further aging.

General Rules of Procedure

The time intervals normally used for the grid step and the plate step vary widely, depending upon tube materials, construction, and application, and the condition of the tube after exhaust. Although it is not practical to specify these time intervals, certain criteria may be given. The time for the grid step is usually limited by the conditions mentioned previously. Because the plate step is not as severe with respect to the cathode, the time may be considerably longer than that of the grid step. Time for the grid step may be 5 to 20 minutes and for the plate step, 10 to 60 minutes. It is first nec-

essary to determine electrode voltages for proper dissipation levels, and then to observe the changing characteristics until appropriate time intervals can be selected. The finished, aged tube must have the activation level and stability to satisfy the electrical characteristics at test and application and the life-test requirements.

Summary

Before going on to other considerations the aging procedures already discussed will be summarized. The three steps in the conventional aging schedule have been presented as the hotshot, grid step, and plate step, as shown below.

Procedure	Step		
	Hot Shot	Grid Step	Plate Step
Heater Voltage — Per Cent of Rated Values	150-200	120-140	120-140
Control-Grid Voltage	Off	On	Off
Screen-Grid Voltage	Off	On	On
Suppressor Grid	Grounded	Grounded	Grounded
Plate Voltage	Off	On	On
Time — Minutes	2-5	5-20	10-60

CIRCUITRY AND EQUIPMENT

The element voltages used in aging should be further defined. The simple case is the direct application of positive dc voltage to the grids and plates with respect to the grounded cathode. Under this condition, there is continuous conduction and element dissipation in the conventional tube construction. For given values of voltage, therefore, dc provides maximum dissipation and maximum aging. Maximum dissipation, however, is not always desirable, because continuous dc aging of fragile control grids, for example, may cause deformation and sublimation of plated or clad materials. To avoid this situation, ac voltage may be applied to the grid. Simultaneous ac grid aging with dc plate aging is often used to give maximum plate aging under ac grid conditions. Tube conditions are shown in Fig. 1.

Although ac aging has been presented as a means to moderate grid aging, it is often necessary to consider complete ac aging for types of ac application, such as certain rectifier-type diodes and other types for use in class C application. High-voltage, short-time pulse aging conditions are also to be considered for types which must serve in pulse applications. While ac plate aging is specifically indicated in the cases mentioned, it may be used for plate aging of conventional types for conventional dc applications. As shown in Fig. 1B, the character of the plate current is controlled essentially by the grid voltage. Because plate current pulsates when ac voltage is applied to the grid, the plate voltage

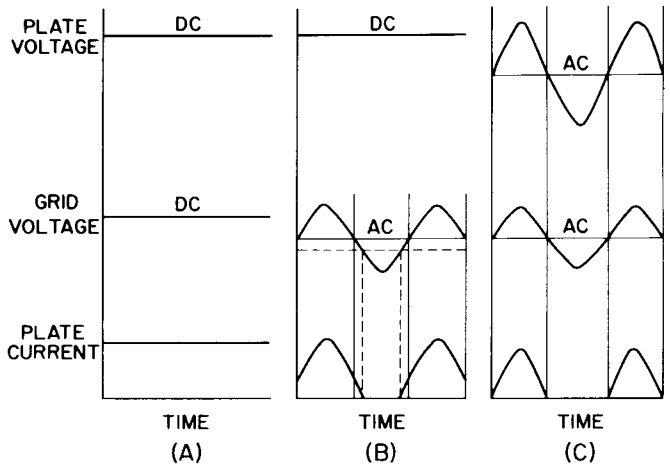


Figure 1. Basic AC and DC Voltage and Current Characteristics

may also be ac. There will be little change in the shape of the plate-current characteristic and a homogeneous supply-voltage system can be used. When ac voltage is used on both grid and plate, it is necessary to adjust element voltages and aging time to obtain dissipation comparable to dc conditions. Alternating supplies and phase differences between the supplies of the conventional three-phase system make the pulsating results considerably different in their dissipation qualities from those of continuous dc aging.

When fixed supplies are used, the conventional filament-type lamp is usually used to adjust and control element voltages. The lamps, which may be used in series with each element or each supply line for each tube on the aging system, provide a simple means of varying element voltages, and also serve as current-limiting loads. The lamps also act as indicators of defective tubes and sockets. An even more flexible system uses continuously variable voltage supplies in conjunction with a lamp load. Element voltages are changed by simple adjustment of the supplies, but the lamp remains to serve its protective and indicating functions.

Basic Circuits

Figs. 2 to 6 illustrate a number of arrangements needed to age several of the common tube constructions with lamps in series with each element.

If aging equipment is to be capable of complete control of any tube, five aging circuits are required. Although this extreme flexibility is sometimes required during tube development, it is not practical for production because of the physical size of the equipment that would be required.

The most generally used type of production aging equipment is the stationary rack consisting of several (usually four) individual panels. The panels for both 7-pin-miniature and octal-tube types have been standardized at 80 sockets, and for the 9-pin-miniature types, at 51 sockets. Four hundred and 255 lamp sockets would be required per panel, respectively, if the flexibility of five aging circuits were to be provided.

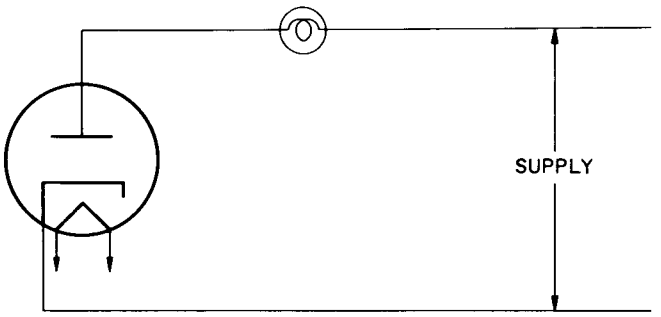


Figure 2. Diode Aging Circuit

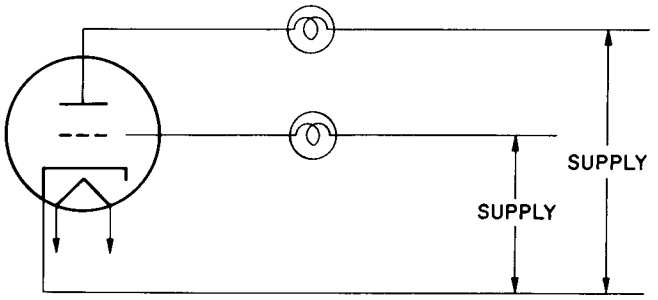


Figure 3. Triode Aging Circuit

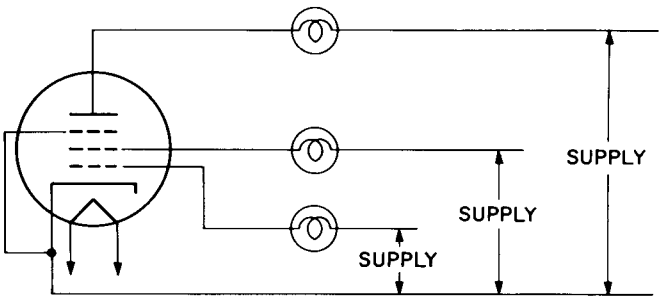


Figure 4. Pentode Aging Circuit

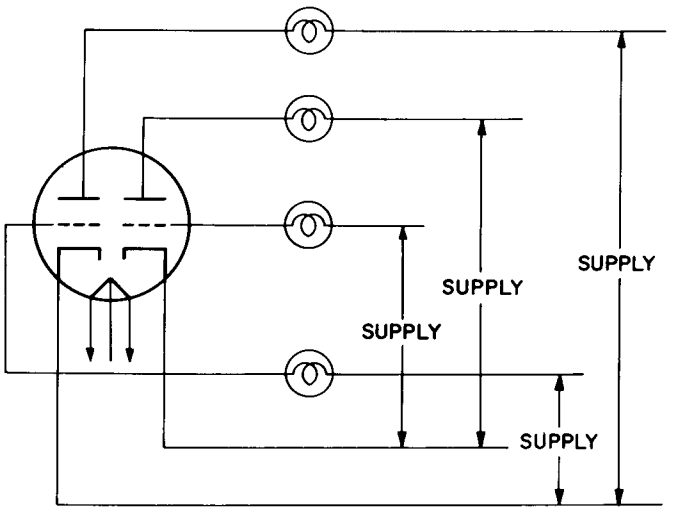


Figure 5. Twin-Triode Aging Circuit

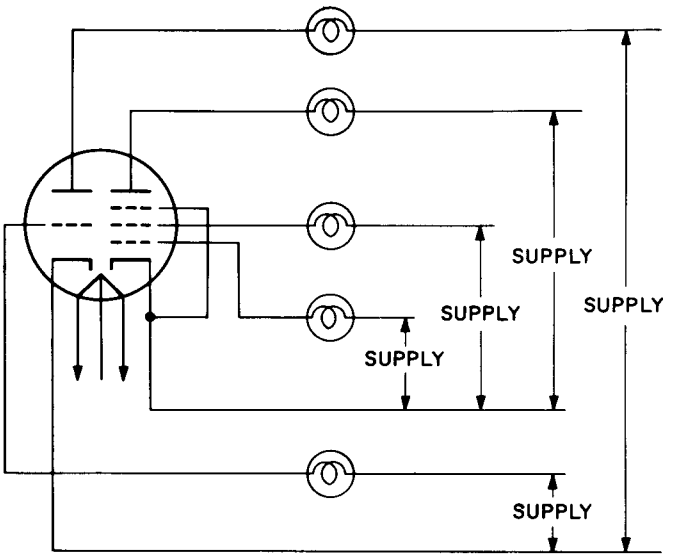


Figure 6. Triode-Pentode Aging Circuit

Four-panel racks would, in turn, involve 1600 or 1275 lamp sockets — unwieldy if not prohibitive quantities. Fortunately, slight changes in aging procedure—changes that are entirely within keeping with ultimate aging requirements — make possible the use of equipment with only two aging circuits rather than five. It is this two-circuit equipment that is used for most production aging. The circuits shown in Figs. 7 to 10 show how aging may be made to conform to standard two-lamp equipment. It is important to remember, however, that this equipment may not be capable of meeting some unusual aging requirements. The most notable examples of tubes requiring unusual aging procedures are in the triode—pentode family. Production equipment has been constructed to provide three-lamp aging, as shown in Fig. 11, for these tube types.

The two-lamp, four-panel stationary rack, which has been described as an example of production equipment, is only one of the types in current use. Automatic equipment of the indexing-conveyor type, using both single-socket and multi-socket panels, is coming into increased use. This equipment accomplishes the same basic aging operations as the manual equipment, but

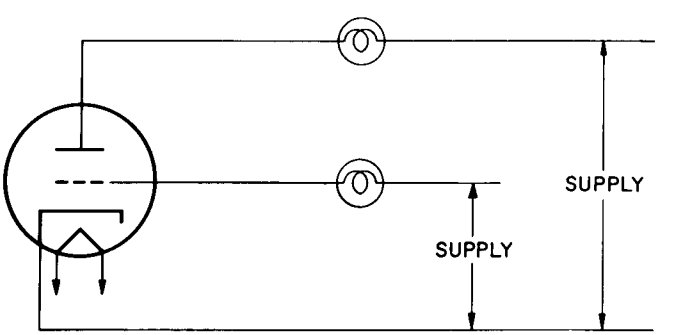


Figure 7. Two-Lamp Triode Aging Circuit

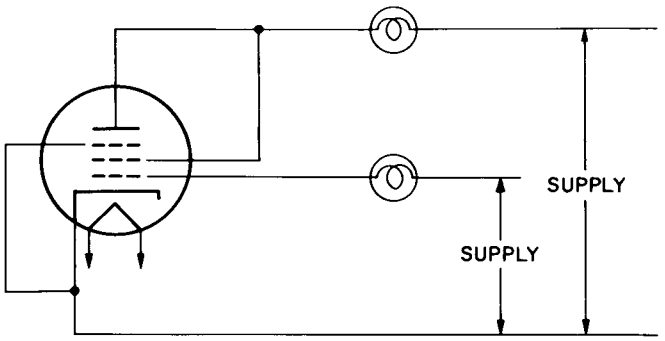


Figure 8. Two-Lamp Pentode Aging Circuit

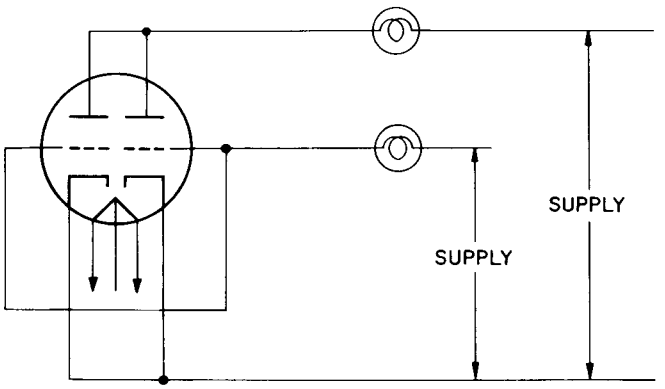


Figure 9. Two-Lamp Twin-Triode Aging Circuit

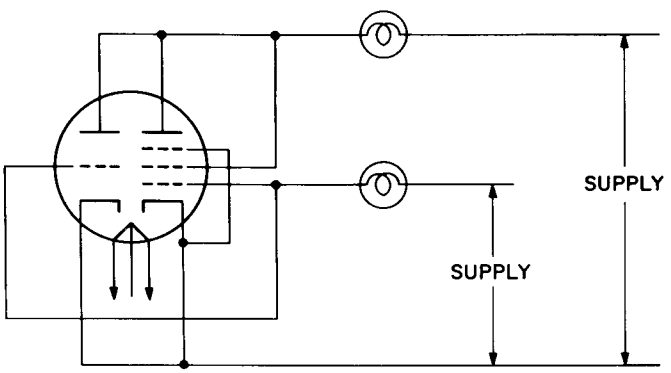


Figure 10. Two-Lamp Triode-Pentode Aging Circuit

increases production rates and removes the need for manual timing and switching.

The relationship between the heater and cathode has not previously been discussed. Although these two elements are normally in physical contact with each other, the heater insulation must isolate the heater wire from the inner wall of the cathode. To determine the insulation properties of the heater coating at operating temperature, voltage is normally applied to the heater-cathode circuit during one or more of the aging steps. Except for unusual applications, the same fixed value of voltage is usually used throughout the aging schedule. In most aging equipment, bus lines serve the entire panel of cathode and heater leads, and the heater-cathode voltage is simply applied to the appropriate pair of bus lines. A satisfactory lot of tubes represents a

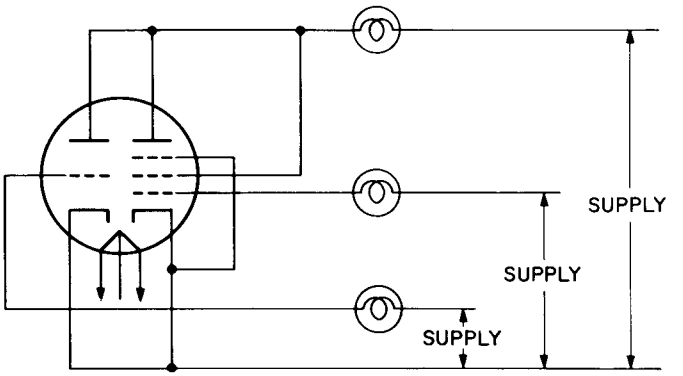


Figure 11. Three-Lamp Triode-Pentode Aging Circuit

no-load condition, and therefore, only one series lamp per rack is used to provide protection and indication in the event of a heater-cathode short. A low-current lamp, such as a small neon lamp, is normally used as a pilot in parallel with the supply. (See Fig. 12).

The most common variation from ac heater-cathode aging occurs in the aging of the damper diode which normally supports high unidirectional heater-cathode voltage in tube operation. In this case it is normal to use high-voltage dc aging (with the heater negative with respect to the cathode). Similar pilot and protective-indicating lamps may be used. Special aging equipment is also used to provide heater-cathode conditions more nearly simulating actual tube operating conditions. One type of special equipment is the "spot-knocker" which is used to operate on both heater-cathode and plate-cathode circuits. Spot-knocking is used for damper tubes and other high-voltage types to duplicate frequencies and voltages encountered in certain tube application. Special equipment is needed for accomplishing this operation because it is not feasible to incorporate unusual circuitry into basic aging equipment. The term spot-knocking suggests that heater-cathode aging serves a functional purpose rather than only an indicating purpose, and this is true. High spots of heater coating (and occasional foreign material on the inner wall of the cathode) are areas where high-voltage arcing or breakdown may occur. Spot knocking is an attempt to remove these areas of trouble during tube manufacture, rather than to allow them to affect circuitry in the tube's subsequent application.

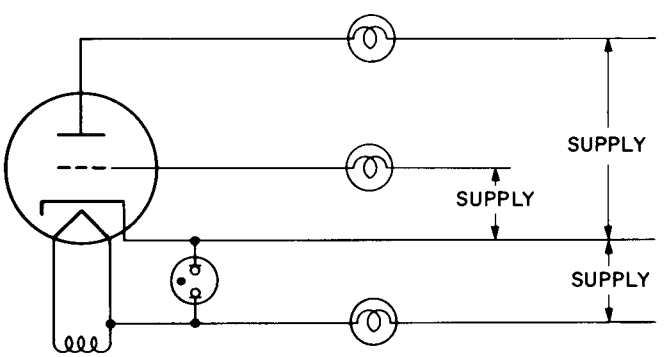


Figure 12. Triode Aging Circuit Showing Heater-Cathode Aging

The heater-cathode aging operation, which was presented simply as an indicating device, actually reduces heater-cathode leakage and hum by reducing or eliminating leakage-producing contaminants and emission material in the heater-cathode system. During hotshot, when the highest processing temperatures are encountered, a thermal effect helps vaporize heater-coating surface contaminants. In aging steps that follow when the heater temperature is lowered, application of heater-cathode voltage introduces an electrolytic effect which causes migration of contaminants to free surfaces from which they are later evaporated. Ac heater-cathode voltage has proven to be more generally useful in receiving tubes than dc for this purpose, since ac reduces the heater-cathode leakage in both directions of conduction. However, the special case which requires dc heater-cathode aging considers only the direction in which the tube will operate — heater negative with respect to the cathode in the case of the damper diode.

The complete two-lamp aging circuits actually include heater-cathode arrangements as shown in Figs. 13 and 14. The triode circuit (Fig. 13), is given as a simple general type using ac heater-cathode aging; the damper diode circuit (Fig. 14), is presented as an example of dc heater-cathode aging.

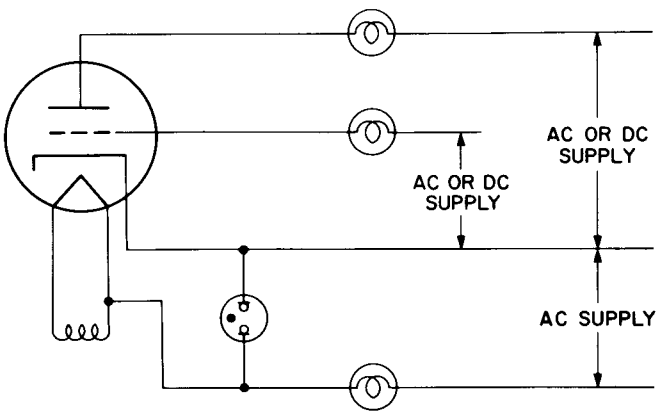


Figure 13. Triode Aging Circuit Showing AC Heater-Cathode Aging

VARIATIONS OF TREATMENTS AND SPECIAL PROCESSING

Elimination of Grid Step

The three-step aging schedule is a basic aging approach, but many variations exist to accomodate different tube types. The most common deviation is the elimination of the grid step. When a tube is processed with little contamination of the grid during exhaust and hotshot, and hotshot and plate aging are sufficient for good activation, it may not be necessary to age the grid for clean-up if the type does not have grid emission problems.

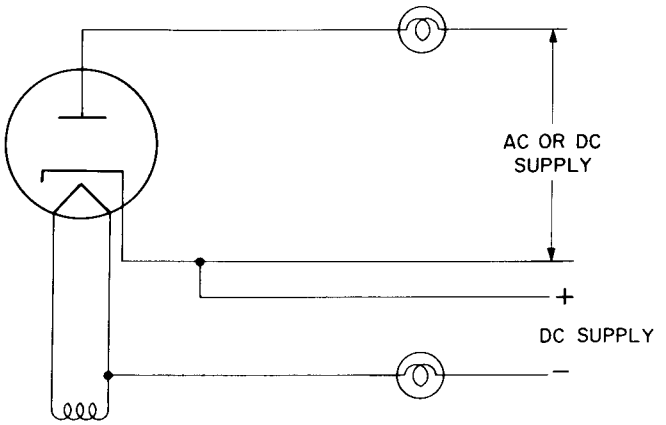


Figure 14. Damper Diode Aging Circuit Showing DC Heater-Cathode Aging

Special Cycling

Another process contains two or more cycles of the basic three-step schedule. This variation is used when one cycle does not result in adequate activation and degassing. Excessive parts degassing after hotshot may poison the cathode and necessitate a repeated cycle of hotshot and the subsequent aging steps. This approach is widely used when processing temperatures are limited in an effort to suppress contamination and the formation of leakage paths. A common adjunct to the multicycle aging process is the cooling step between cycles. Although this has not been completely evaluated, it appears that the cooling period allows the degassing process to be delayed while gas may be adsorbed at the cold nonconducting surfaces. Cooling is also used in certain types which are processed with unusually high short-time heater voltages for elimination of hum. Because high temperature is required to reduce heater contaminants or heater emission (from stray cathode or getter material), but is not intended to produce high cathode temperatures, cooling steps are inserted to process the heater without excessively heating the cathode.

Problem Tube Types

The relation of aging to exhaust influences the variations in aging schedules. The exhaust processing system is often incapable of giving optimum breakdown and degassing. These limitations are related primarily to tube construction. Optimum cathode breakdown is unobtainable in short mounts because of the nonuniform cathode heating characteristics that result from end-cooling of the cathode. Close-spaced diodes, such as damper tubes, may also have nonuniform cathode heating during exhaust. This nonuniformity results from heating of the cathode by radiation from the anode during rf heating of the anode. Another difficulty is introduced with the dual-section tube type which includes nonuniform cathodes and/or heaters. Uniform breakdown is extremely difficult with this combination of parts, and the exhaust operation usually cannot deliver a tube to aging with balanced cathode-breakdown conditions. Parts degassing is also limited by tube construction. Rf pickup loops in tube parts may not be located in the area of the part which requires the heating. Outer parts

that run cold in operation may shield inner parts that require degassing. Critical grid materials may limit over-all heating characteristics.

Contact Potential Control

Contact potential* is a problem in some tube types. The aging process can be designed to minimize this problem. Contact potential results from the differences in work functions of the cathode and adjacent elements. To minimize contact potential, the work functions of the adjacent elements should be as close as possible to that of the cathode.

Two approaches to minimization of contact potential at aging have been used. The first seeks to avoid the increase in work function that results when foreign materials are deposited on tube elements; the second seeks to reduce the work function of the tube elements by deliberately causing the deposition of cathode coating on adjacent parts.

Tube elements may be kept free of foreign materials during aging by the use of a long, low-temperature aging schedule.

Intentional deposition of cathode coating is obtained by the use of a severe hotshot at the end of the aging process. Although deposition of cathode material may successfully reduce contact potential, it is hazardous. Deposition of cathode coating material on an adjacent grid to reduce the effective work function, also improves the emission properties of the grid. Great care must be taken, then, that contact potential is not reduced at the expense of increased grid emission.

Parasitics

Aging equipment which is wired to process large quantities of tubes often introduces unusual oscillatory effects. Parasitic circuit elements are established due to the extensive wiring and these may cause a rack to develop oscillation with any indeterminate number of tubes being processed. Tubes with transconductance values of roughly 5000 micromhos and higher have a tendency to oscillate. The incidence of oscillation is unpredictable with respect to the number of tubes or their locations on the rack. This oscillation produces aging conditions and variations which are as unpredictable and uncontrollable as the phenomenon itself. A possible method of eliminating this effect is the introduction of a capacitance in each tube socket from plate to cathode or from grid to cathode. The capacitors used should be small noninductive elements, such as mica capacitors of 500 to 1000 micromicrofarads. This addition should be resorted to when oscillation develops on a given rack with a given tube type. However, a more positive approach might be the addition of such capacitors at the time a rack is built. Although this would be expensive on multisocket aging equipment, it would be an extremely valuable control on oscillation.

* See article on "Contact Potential and Grid Currents," by E. R. Schrader.

STABILIZATION

AREAS OF APPLICATION

While conventional aging is suited to high-speed mobile manufacturing operations and normally provides adequate tube processing for conventional application, further refinement in processing is sometimes necessary. Tubes for many industrial applications, notably Premium tubes, must deliver unusually stable, long-life, trouble-free operation at a high quality level. Applications in these areas depend upon the receiving tube as a reliable system element; the consequences of tube variation or failure may be costly in terms of expensive equipment or process failure. Even more significant is the military application where receiving tube reliability is vital.

Passive-Type Cathodes

To assure stable long-life operation, tube design normally calls for materials and processes which provide these characteristics. The fact that passive cathode materials are used more in this area than in tubes made for entertainment application is of fundamental importance to a discussion of aging and associated processing. These materials are used to suppress the development of interelectrode leakage and cathode interface, both of which become increasingly serious with long-life operation.

In addition to material considerations, lower cathode temperatures are maintained during operation to further suppress the formation of leakage paths. Because activation of cathode coating during aging depends to a large degree upon the amount of reducing elements in the cathode alloy, tubes made with passive cathode materials require more careful and extended aging for optimum stability. The time interval feasible for manufacturing mobility is not normally sufficient for complete processing of industrial tube types, and it is therefore necessary to prolong the processing on a separate basis. (It is not suggested that the stabilization operation be used to compensate for improper aging procedures; stabilization provides for an extended period of operation to produce subtle changes under operating conditions.) Stabilization is actually carried out under electrical conditions closely related to life-test conditions so that the tube may be subjected at this point to conditions it will see in its application. If this process is properly carried out, the tube undergoes its minor early-hour variations while still under manufacturing control. These early-hour variations are normal in most receiving tubes, but are not usually objectionable in many applications. However, in critical industrial applications which depend upon an extremely stable circuit element, these early-hour variations must be eliminated or taken to completion. Stabilization, then by continued tube processing under life-test or operating conditions, provides for refinements in activation and parts degassing and the consequent variations in conduction under these conditions, to be completed. (The exceptional tube type which displays no variation at stabilization represents an optimum processing condition. Stabilization for this tube type becomes an instrument for monitoring production processing success. This

is perhaps the ultimate method for assuring maximum quality level for outgoing product.) The time that has been found necessary for stabilization of tube operation or tube characteristics, has varied from several hours to several days, but a practical limit for most types has been established at 48 hours.

Elimination of Potential Inoperatives

Another consideration related to tube stabilization is its use as a "testing method," rather than only as a processing function. Catastrophic tube failure is ordinarily caused by shorted or open tube elements, either of which renders the tube completely inoperative. These inoperatives normally occur during the first several hours. Tube stabilization provides an excellent opportunity to find early-hour inoperatives, and because stabilization usually continues beyond the apparent critical time, the finished product may reach a high quality level in terms of subsequent inoperatives.

CIRCUITS AND EQUIPMENT

Stabilization requires extensive physical equipment, containing a great number of sockets, because long-period stationary processing of this type obviously is not consistent with the speeds of the other manufacturing operations. Early equipment designed for stabilization did not provide for the application of specific element voltages at the tube sockets. The important consideration in this design was to effect life test dissipations at the elements by any reasonable or convenient means. The equipment was designed to operate ten tubes in parallel in the manner shown in Fig. 15 (triodes are shown for the sake of clarity).

The heaters and heater supply are not shown because they are normally conventional variable ac supplies. Heater-cathode conditions are not shown, but isolated variable ac or dc supplies are normally used, and, as in aging equipment, are simply connected to the cathode and heater bus lines. The dc plate voltage supply shown is used for most types. There are a limited number of dual-diodes which see ac service and which, accordingly, are stabilized with ac plate voltage. The dual-diodes also have appropriate circuit elements which differ from the standard dc stabilization circuit shown.

The circuit uses a conventional filament-type lamp in series with the cathodes to develop the grid bias needed to attain the required plate dissipation; the cathodes are actually at a voltage above ground equivalent to the voltage necessary for proper grid bias. The lamp is selected to produce the desired voltage drop with total cathode current from ten tubes. Although the actual plate voltage with respect to the cathode may be considerably different from the specified life test value, the grid bias can be adjusted to give the necessary plate current for equivalent dissipation.

The grid resistor R_g helps duplicate normal tube application by allowing the tube to react in the same manner that it would in use. The grid resistor may subject gassy tubes to a destructive condition. If gas is considered as a negative grid current, the grid resistor, which is usually of high value (from 0.1 to 1.0

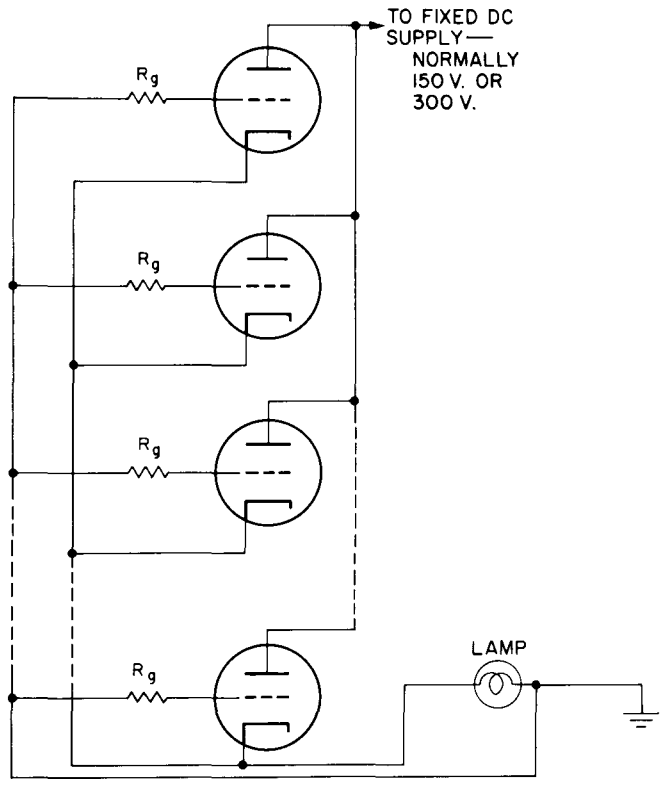


Figure 15. Ten-Tube Stabilization Circuit Showing Lamp Bias

megohms), effectively becomes an instrument for decreasing the grid bias when grid current flows. Since this reduction results in increased plate current, and in turn, increased gas ionization, the cycle may proceed to the tube's destruction. In this manner, tube stabilization adds another function in improving the eventual quality level of finished product.

In present equipment, the 10-socket unit is an integral part of a 50-socket panel of five such rows of sockets. The full standard rack consists of 32 panels, or 1700 tube sockets and 160 lamp sockets for the complete rack.

The equipment previously described, although it provides a rugged, flexible process, is subject to severe limitations. First, it may be difficult to obtain precise values of dissipation with available lamps. Since bias voltage is normally critical in determining plate current, the stepwise variation of bias obtainable by changing lamps may not be sufficiently precise. It may even be necessary to remove one or two tubes from each row of sockets to arrive at proper conditions on the remaining tubes, or as an alternative, the tubes may be operated at undesirable dissipation conditions.

An improved method employs individual cathode bias resistors as shown in Fig. 16. Each tube (or section in dual-section types) assumes a grid bias condition which is essentially independent of other tubes on the rack, and which may be established accurately due to the general availability of a wide range of resistor values. The lamp, which formerly served as a biasing element, is used to regulate plate voltage with respect to the cathode, since the voltage drop across the lamp

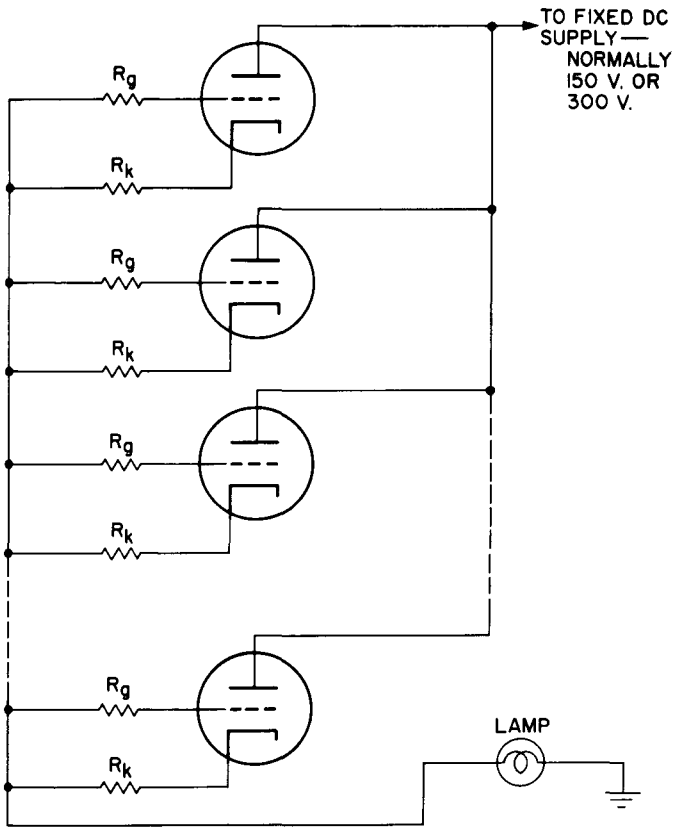


Figure 16. Ten-Tube Stabilization Circuit Showing Cathode Bias Resistors

establishes the plate-to-cathode voltage. Since the total current (from ten tubes at given bias and plate voltage conditions) and the required voltage drop are known, the lamp size may be selected from available curves.

The improved stabilization method described does not represent any basic equipment change. The additional resistors for self-biasing are simply added at the sockets of the conventional 50-socket panel, and the grid bus is connected to the correct side of the lamp. Although the lamps serve different purposes, the physical equipment for either method is essentially the same.

Individual tube self-bias provides for closer self-regulation than the ten-tube lamp-bias method described earlier. But use of the lamp as a voltage-dropping element still results in some tube interdependence. To overcome this interdependence, individual plate resistors have been used to drop supply voltage to correct plate voltage. The resistor power ratings required for safe operation are excessive, and the method has not been used extensively.

The latest stabilization equipment contains variable dc supplies for positive elements and also for grid bias. The choice between fixed-bias and self-bias is determined by the circuit requirements of the tube type.

SUMMARY

FACTORS AFFECTING AGING AND STABILIZATION SCHEDULES

There are a prohibitive number of variables to consider in an attempt to specify precise aging and stabilization schedules beyond the basic approach described. Aging procedure has often been based upon the experience of the designer and the many variables peculiar to a tube type. The following review of pertinent variables and the considerations they introduce, indicate why it has not been feasible to specify a ready "formula" for tube aging:

Relation of Aging to Exhaust

1. Degree of cathode breakdown from carbonates to oxides. (Determines the amount of subsequent activation necessary at aging.)
2. Temperature of heater and cathode during exhaust and the limitations imposed with respect to new degassing during aging.
3. Parts degassing efficiency during exhaust. Construction variations, such as parts design, with respect to rf pickup, affect degassing properties and limitations during exhaust.
   a. Plate and outer metal parts design provide wide range of shielding characteristics.
   b. Getter location may limit rf coupling for best degassing of getter and other parts.
   c. Grids may be of materials which limit over-all rf heating properties of tube.

Tube Parts and Materials

1. Heater limitations to temperatures desired.
2. Cathode sleeve material.
   a. Active materials—alloying agents which sublime readily limit aging temperatures.
   b. Passive materials require extended aging for activation.
3. Cathode-coating material. Dense materials do not vaporize onto adjacent tube parts as readily as fluffy dry materials, but require heavier aging.
4. Grid material. Surface materials limit grid aging by their sublimation properties.
5. Plate materials. Degassing determined to a large degree by surface character—spongy carbon coating tends to be more gassy than smooth-surfaced material.

Tube Design

1. Cooling properties.
   a. Mass of parts and general construction determine radiation properties and effective temperatures.
   b. Mica fit of parts, particularly of cathode, determines cooling and end effects which influence cathode temperature during processing.
2. Electrical connections. Common element connections provide limitations in aging variations.

IMPORTANCE OF MEASUREMENTS

Formulation of aging schedules should be accompan-

ied by accurate electrical measurements. Aging voltages may be varied through lamp selection with fixed supplies, or through variable supply values, with periodic measurement of element currents and dissipations. This should be supplemented by periodic measurement of emission and gas to determine the resultant activation and clean-up. Of course, this must be followed by complete electrical testing and life test.

For dc conditions, dissipation may be determined by use of a dc voltmeter and dc ammeter, and the simple product of the resultant pair of values. Ac conditions present a different problem. Because of possible phase differences between supplies and because of the conduction characteristics, the wave shapes of element voltages and currents are not the pure sinusoidal shapes which the normal wattmeter can measure. The ac plate and/or grid voltage curve may be compressed in the area of conduction because of a voltage drop that occurs in the series lamp when tube conduction occurs. The element current curve may be compressed at the peaks in tube types which reach saturation at a voltage below the peak voltage of the ac element voltage applied. Due to the complex nature of these wave shapes, special measuring methods or equipment must be considered.

There is commercially available a special low-power-factor wattmeter, frequency-compensated to 2500 cycles. This instrument approaches a true reading since most of the major frequency components of the complex wave shapes are included in the range of frequencies that the meter is capable of handling. Another instrument which has been used is the thermocouple-type wattmeter, which contains essentially resistive elements (rather than the normal inductive elements) and hence has good high-frequency capabilities. This instrument, however, is severely subject to de-calibration upon overload. The special 2500 cycle meter described above does not have this limitation and, more important, probably is of sufficient accuracy for measurement of the wave shapes with which we are concerned.

An empirical method of limited use has been the measurement of bulb temperature under easily measurable dc conditions and subsequent duplication of the temperature using ac conditions. This is not useful for multi-element dissipation since the individual dissipations are obviously not separable in the bulb temperature reading.

BIBLIOGRAPHY

Herman, G., and S. Wagener, The Oxide-Coated Cathode (2 volumes), Chapman and Hall, 1951

Kohl, W. H., Materials and Techniques for Electron Tubes, Reinhold, 1960

Spangenberg, K. R., Vacuum Tubes, McGraw-Hill, 1948

Diodes

A. P. Kauzmann

Harrison

The diode has the least complicated structure in the family of thermionic tubes. It performs most circuit applications as a simple polarized switch, closing whenever the anode (or plate) is more positive than the cathode and opening when this polarity is reversed. Although this simple concept is useful for an approximate analysis of circuit performance, it must be expanded into more practical form by the tube designer; his switch is not ideal and consequently does affect the circuit performance. The magnitudes of the currents, the voltages, and the dissipations present at the diode terminals are of fundamental importance to the tube designer. This information is needed to determine the physical parameters for a particular structure and to fix the electrical ratings upon a finished tube design.

A considerable portion of this article must therefore deal with conventional circuits using thermionic diodes. The quantitative analysis of these circuits will demonstrate the effects of the diode on circuit performance and, vice versa, the effects of the circuit on the diode performance. The limiting qualifications of the diode can be determined from the latter.

DIODE SYMBOLS

Solid-state diodes and thermionic diodes can perform the same function in any circuit. Their symbols are compared in Fig. 1. In the solid-state device, one electrode is usually marked positive (see Fig. 1A). The positive electrode corresponds to the cathode of the thermionic diode. Note that for any two-terminal load, either in series or in parallel with the rectifier device, the end of the load closest to the positive-marked electrode will also be positive. The arrow indicates the direction in which current, in the conventional sense, flows during conduction; its direction is always in the opposite sense of electron flow.

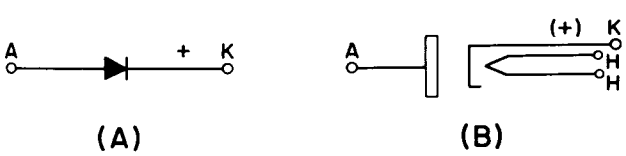


Figure 1. Conventional Symbols: (A) the solid-state rectifier device, (B) the thermionic-emitting diode

DIODE THEORY

Plane Parallel Electrodes

Neglecting the effects of initial electron velocities on the diode characteristic by the use of the Langmuir equations[1] (which do neglect this effect), greatly simplifies the diode working formulas. This procedure is justifiable in most cases, because the initial electron velocity effects are trivial at positive anode-to-cathode voltage in excess of only a few volts. Then, the diode current i for a plane-parallel cathode and anode is given by

$$i = \frac{2.334A}{d^2} e_d^{3/2} = Ge_d^{3/2} \text{ microamperes} \quad (1)$$

where A = cathode area
d = cathode-to-anode spacing
e_d = voltage between cathode and anode in volts (corrected for contact potential)
G = diode "perveance"

In Eq. (1) and the following Eq. (2), no units are designated for expressing A, d, L, and r; they can be expressed in any units provided they are consistent for numerator and denominator.

Concentric Electrodes

For cylindrical electrodes where the cathode is inside and concentric with the anode and with the same assumptions as above,

$$i = \frac{14.66\ L}{r\beta^2} e_d^{3/2} = Ge_d^{3/2} \text{ microamperes} \quad (2)$$

where L = cathode length
r = anode inside radius
β^2 = $f(r/r_o)$, where r_o is the cathode radius
e_d = voltage between cathode and anode in volts (corrected for contact potential)
G = diode "perveance"

The values of β^2 and β have been evaluated by I. Langmuir;[1] however, in the chart (Fig. 12 of the article by L. Scholz) the perveance G is plotted as a function of r and r_o for a unit length of one millimeter. This chart inherently contains the factor β^2.

Perveance

The concept of perveance as applied to a diode is

unique because it is dimensionally the quotient of current divided by the three-halves power of voltage, or $I/V^{3/2}$. This perveance is the G factor of Eqs. (1) and (2) indicating that, besides a numeric constant, the perveance contains all the geometric quantities determining current flow. A high-perveance diode implies either close electrode spacings, a large cathode area, or a combination of both as compared to a normal-perveance diode. It is expedient to note that, at a diode plate voltage of one volt positive, the magnitude of plate current equals the magnitude of the diode perveance. At positive plate voltages of 4.65 or 21.5 volts, the magnitudes of the plate current are 10 times and 100 times the magnitude of the tube perveance, respectively.

When the emitter is in the form of a V-shaped or M-shaped filament, the effective cathode area component of the diode perveance is determinable through a suggestion by Kusunose.[2] As shown in Fig. 2, this effective area consists of both sides of an imaginary flat strip lying in the plane of the filament. The width of this strip is twice the distance between the filament center line and the plate, and its thickness is assumed to be zero.

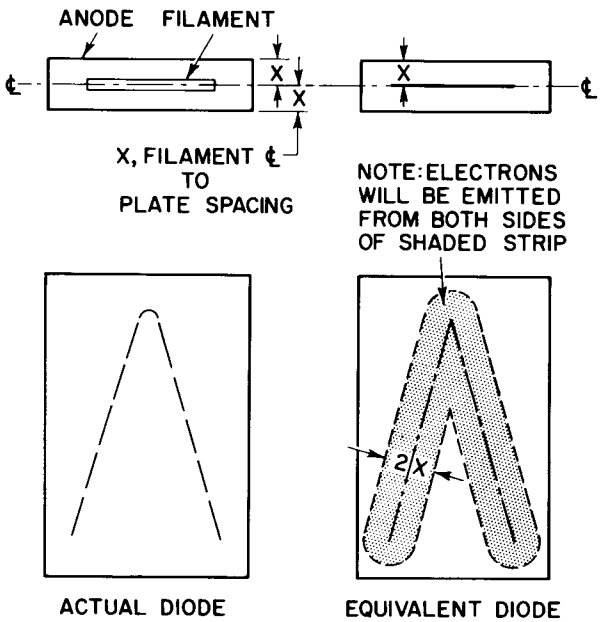


Figure 2. Effective Cathode Area of a Filament-Type Diode

Diode-Plate Characteristic

In Fig. 3, diode-plate characteristics are plotted on log-log graph paper as straight lines with the slope of 3/2. If this characteristic is plotted on log-log paper from observed data using the applied E_p as the abscissa, the 3/2 slope should appear at the region of higher voltages. In the 1-to-10 volt positive region, the curve will usually pull away, either above or below the straight line. Inspection will show that the addition or subtraction of a small constant voltage of the order of + 1 volt will shift the curved part of the characteristic to the 3/2 slope line of the higher voltage. The magnitude of this voltage including its positive or negative sign is the so-called contact potential, ϵ. Strictly speaking, it also includes some initial-velocity effect. Below approximately plus one volt, the effect of initial velocity predominates over space charge effects and the 3/2 power law usually no longer holds. (See the article by L. Scholz, where the effect of initial electron velocities on current flow is presented.)

Dynamic and Effective Diode Resistances

Two resistance concepts are related to the diode anode characteristic. The dynamic diode resistance r_{dyn} is obtained from the differentiation of Eq. (1) or Eq. (2),

$$r_{dyn} = \frac{de_d}{di} = \frac{2}{3} \frac{1}{G^{2/3} i^{1/3}} = \frac{2}{3} \frac{1}{Ge_d^{1/2}} \qquad (3)$$

This resistance would be used in analyzing circuits in which small ac voltages or currents appear across or pass through the electrodes of the diode, respectively. For this case, the diode would be biased with a positive dc voltage which is large compared to the ac voltage applied.

When the diode is acting as a switch (as in the ac-dc power conversion circuits) a second resistance concept is more practical; this effective diode resistance r_d is the quotient of a particular voltage across the diode divided by the total corresponding diode current. For any operating point on the diode characteristic, the diode can be replaced by a simple linear resistor. Then the same voltage and current will be observed across and through the resistor as appeared respectively for the diode. Manipulation of Eq. (1) or Eq. (2) gives the following expression for the effective diode resistance

$$r_d = \frac{1}{G^{2/3} i^{1/3}} = \frac{1}{Ge_d^{1/2}} \qquad (4)$$

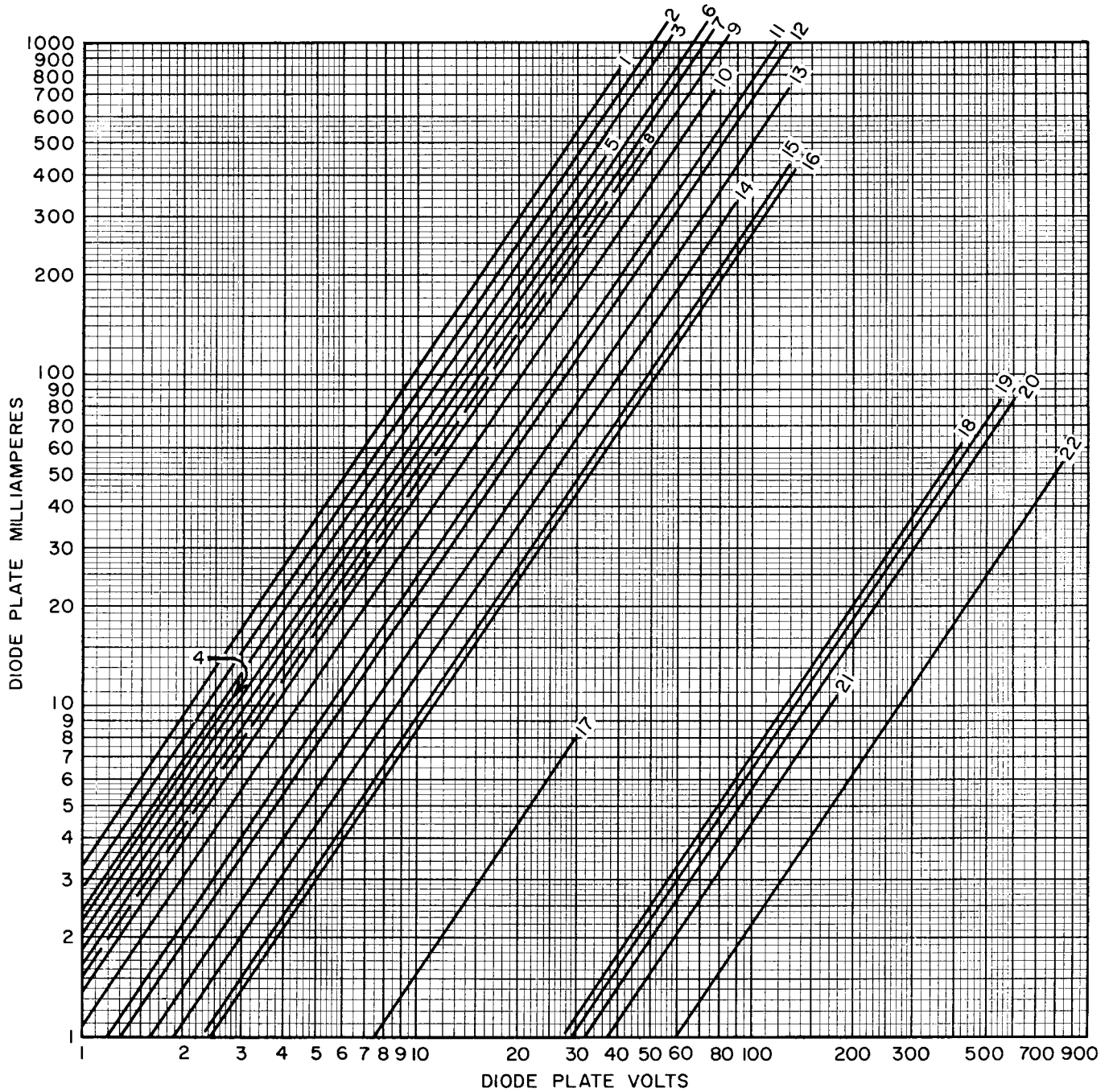
Eqs. (3) and (4) show that both the dynamic and equivalent diode resistances vary inversely with the cube root of the diode current, or inversely with the square root of the diode voltage; furthermore, at the same operating points they bear the simple relationship

$$r_d = \frac{3}{2} r_{dyn} \qquad (5)$$

In Fig. 4 the graphical relationship between a diode characteristic, its dynamic resistance, and its equivalent diode resistance is indicated for one point on the diode curve.

Diode With Added Series Resistance

A constant resistance in series with the diode is almost always found in practical circuits. A qualitative current-vs.-voltage characteristic of such combinations is shown in the small diagram of Fig. 7A. As the series resistance R_s is increased from zero to large values, the current-voltage characteristic changes from a curve (3/2-power) to a straight line. For an analytical solution of this combination, note that the equivalent diode resistance r_d is not a constant. It is therefore expedient to pick a particular set of operating conditions to which current, voltage, and equivalent resistances are normalized. Such a point could con-



TUBE	CURVE	TUBE	CURVE	TUBE	CURVE	TUBE	CURVE	TUBE	CURVE
1A3	17	3A3	20	5Z4	9	6CA4	3	17H3	9
1B3	19	5AS4A	11	6AF3	5	6DA4	3	25Z6	8
1G3/1B3	19	5T4	10	6AL5	6	6DE4	6	35W4	2
1J3	22	5U4G	13	6AU4GTA	2	6H6	14	35Z5GT	2
1K3	22	5U4GB	11	6AX4GT	9	6W4GT	2	36AM3	7
1V2	21	5V3	10	6AX5GT	15	6X4	12	84/6Z4	12
1X2A	19	5V4GA	9	6BC7	1	6X5GT	12	117Z3	17
1X2B	18	5Y3GT	16	6BJ7	4	6Z4	12		
2EN5	6	5Y4GT	16	6BW4	14	12D4	2		

Figure 3. Average Plate Characteristics of Some RCA Rectifier Tubes

veniently (although not necessarily) be chosen for the maximum peak current which is expected to flow. In Fig. 5, let such a reference current be i_1 with the dc reference voltage e_1 applied. Then

$$e_1 = i_1 (R_s + r_{d_1}) \tag{6a}$$

and, with the aid of Eq. (4), at any other applied voltage e

$$e = i\left[R_s + r_{d_1}\left(\frac{i_1}{i}\right)^{1/3}\right] \tag{6b}$$

Dividing Eq. (6b) by Eq. (6a) gives

$$\frac{e}{e_1} = \frac{i}{i_1}\left[\frac{1}{1+r_{d_1}/R_s} + \frac{(i_1/i)^{1/3}}{1+R_s/r_{d_1}}\right] \tag{6c}$$

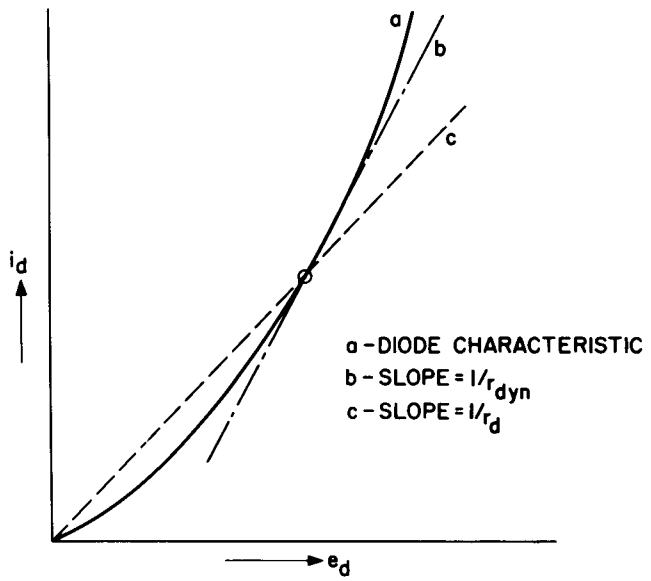


Figure 4. Interrelation of Diode Characteristic, Dynamic Resistance r_{dyn}, and Equivalent Diode Resistance r_d at One Operating Point

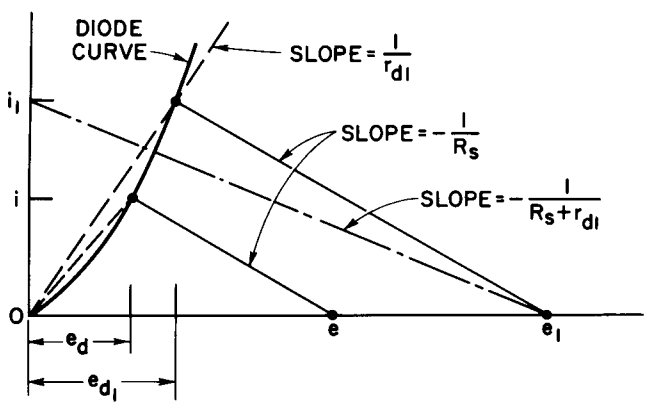


Figure 5. Graphic Determination of Current Flowing Through a Diode with Added Series Resistance R_s

Plotting Eq. (6c) on log-log paper (see Fig. 6) gives a series of practically straight lines intersecting in the reference point. The slope of each straight line is a function of the added series resistance R_s. A new and simpler equation may therefore be written for Eq. (6c)

$$\frac{i}{i_1} = \left(\frac{e}{e_1}\right)^n \tag{7}$$

where the exponent n, a function of R_s/r_{d_1}, represents the slopes of the straight lines of Fig. 6.

In Fig. 7C, the relationship between this slope n and the ratio R_s/r_{d_1} was constructed from graphical solutions of Eq. (6c) similar to Fig. 6.

Substitution of Eq. (6a) into Eq. (7) gives another useful expression

$$i = \left[\frac{e_1^{(1-n)}}{R_s + r_{d_1}}\right] \cdot e^n = Ce^n \tag{8}$$

where C is a constant equal to the bracketed term

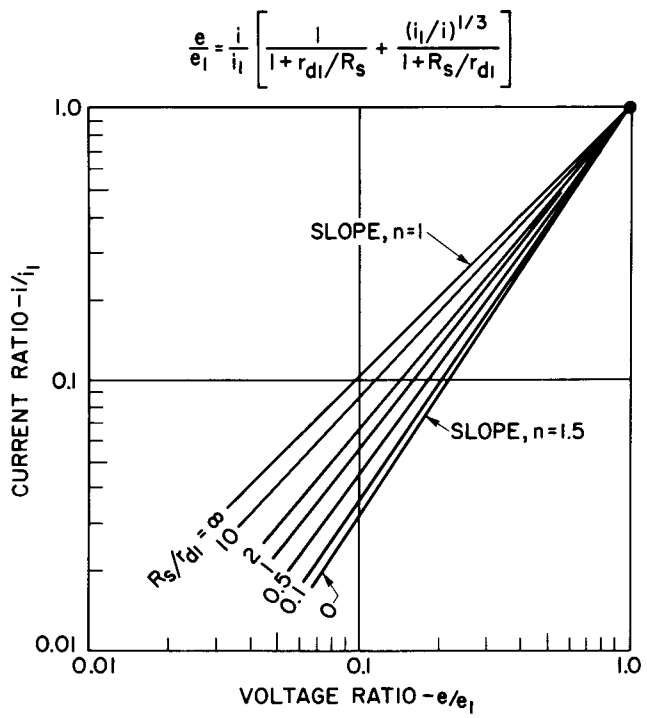


Figure 6. Log-Log Plot of Eq. (6c) for Different Values of R_s/r_{d1}

$$\frac{e}{e_1} = \frac{i}{i_1}\left[\frac{1}{1+r_{d1}/R_s} + \frac{(i\,1/i)^{1/3}}{1+R_s/r_{d1}}\right]$$

The equivalent resistance R at the terminals a, b of the series circuit of Fig. 7B can be deduced in a similar manner giving

$$\frac{R}{R_1} = \left[\frac{i_1}{i}\right]^{\frac{n-1}{n}} = \left[\frac{e_1}{e}\right]^{n-1} \tag{9a}$$

where the value of n in the exponent term is also that shown in the curve of Fig. 7C.

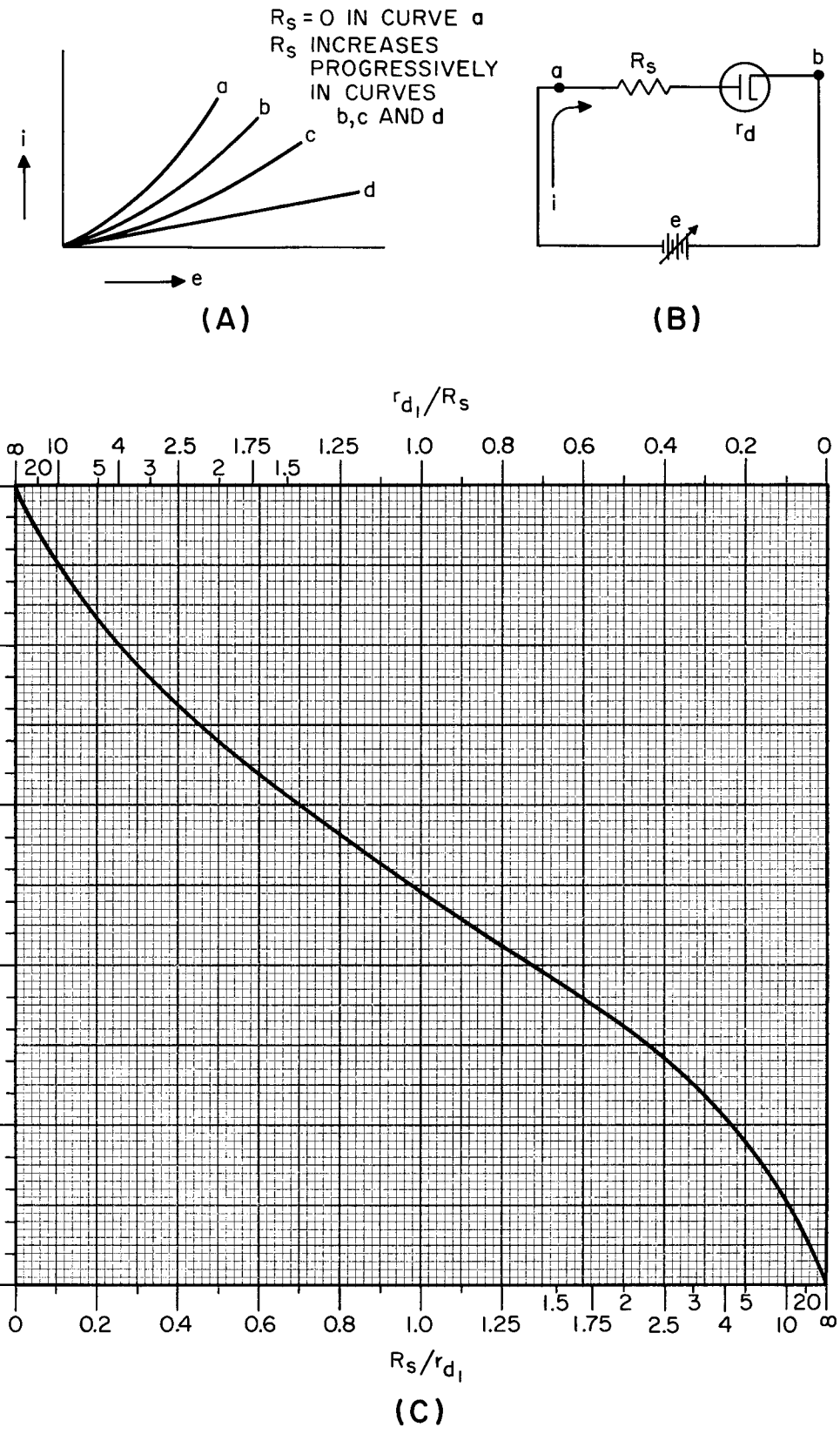


Figure 7. Effect on Diode of Adding Series Resistance R_S: (A) qualitative relation between current and voltage as R_S is increased, (B) circuit defining equivalent resistance R at terminals (a, b), (C) value of exponent n as function of R_S/r_{d1} or r_{d1}/R_S

Substituting $R_1 = R_s + r_{d_1}$ into Eq. (9a) gives the following equations

$$R = \left[(R_s + r_{d_1})\, i_1^{\frac{n-1}{n}} \right] \Big/ i^{\frac{n-1}{n}} = \frac{C'}{i^{\frac{n-1}{n}}} \qquad (9b)$$

$$R = \left[(R_s + r_{d_1})\, e_1^{\,n-1} \right] \Big/ e^{n-1} = \frac{C''}{e^{n-1}} \qquad (9c)$$

where C' and C'' are constants equal to their corresponding bracketed terms. Simply restated, the combination of a series resistance and a diode can be replaced by a two-terminal "black box" having the voltage, current, and resistance relationships expressed in Eqs. (7) and (9a). The combination will appear as a linear resistance as the exponent n approaches unity as a limit, that is, when R_s is very large compared to r_d.

DIODE OPERATION INTO A SHORT-CIRCUITED LOAD

Operation of a diode from a sine-wave plate-voltage supply into a short-circuited load is useful for checking diode emission. The ac-voltage method has an advantage over the dc-voltage method because the diode dissipation is reduced to less than 25 per cent for a given maximum emission current, Eq. (10r). The circuit is simple and consists of an ac supply shunted by the diode. The diodes are assumed to be space charge limited.

The voltage e across the diode is the ac supply voltage and is given by

$$e = \hat{E} \sin\theta = \sqrt{2}\, E_{rms} \sin\theta \qquad (10a)$$

The corresponding diode current i is given by

$$i = Ge^{3/2} = G\left[\hat{E} \sin\theta \right]^{3/2} \qquad (10b)$$

where G is the diode perveance.

The peak current $\hat{I}$ flows when the ac voltage is at its positive peak voltage with respect to the diode cathode and is given by

$$\hat{I} = G\hat{E}^{3/2} \qquad (10c)$$

The dc current $\bar{I}$ is the integral of the current flowing during the half-cycle of conduction, averaged over a full cycle and is given by

$$\bar{I} = \frac{2}{2\pi} \int_0^{\frac{\pi}{2}} i\, d\theta = \frac{\hat{I}}{\pi} \int_0^{\frac{\pi}{2}} \sin^{3/2}\theta \; d\theta \qquad (10d)$$

The rms current I_{rms} is the root-mean-square value of the diode current averaged over a full cycle and is given by

$$I_{rms} = \sqrt{\frac{2}{2\pi} \int_0^{\frac{\pi}{2}} i^2 d\theta} = \frac{\hat{I}}{\sqrt{\pi}} \sqrt{\int_0^{\frac{\pi}{2}} \sin^3\theta \, d\theta} \qquad (10e)$$

The total plate dissipation W_p is obtained by integrating the instantaneous dissipation; it is expressed as the product of the instantaneous diode current and voltage during the conducting half-cycle and is averaged over a full cycle.

$$W_p = \frac{2}{2\pi} \int_0^{\frac{\pi}{2}} ei\, d\theta = \frac{\hat{E}\hat{I}}{\pi} \int_0^{\frac{\pi}{2}} \sin^{5/2}\theta \, d\theta \qquad (10f)$$

The above definite integrals have solutions resolvable with the help of gamma functions since

$$\int_0^{\frac{\pi}{2}} \sin^n \theta \, d\theta = \frac{\sqrt{\pi}}{2} \left[\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}+1\right)} \right] \qquad (10g)$$

Solving Eqs. (10d) through (10f) and substituting the equalities shown in Eqs. (10a), (10b), and (10c) gives the following group of identities for the diode when an ac voltage is applied directly between anode and cathode:

$$\hat{I}/\bar{I} = 3.59 \qquad (10h)$$

$$I_{rms}/\bar{I} = 1.655 \qquad (10i)$$

$$\hat{I}/I_{rms} = 2.17 \qquad (10j)$$

$$\hat{I} = 1.682\; GE_{rms}^{3/2} = G\hat{E}^{3/2} \qquad (10k)$$

$$\bar{I} = 0.469\; GE_{rms}^{3/2} = 0.278\; G\hat{E}^{3/2} \qquad (10l)$$

$$I_{rms} = 0.776\; GE_{rms}^{3/2} = 0.461\; G\hat{E}^{3/2} \qquad (10m)$$

$$W_p = 0.544\; GE_{rms}^{5/2} = 0.229\; G\hat{E}^{5/2} \qquad (10n)$$

$$= 0.702\; I_{rms}\, E_{rms} = 0.496\; I_{rms}\, \hat{E} \qquad (10p)$$

$$= 1.160\; \bar{I}\, E_{rms} = 0.820\; \bar{I}\, \hat{E} \qquad (10q)$$

$$= 0.324\; \hat{I}\, E_{rms} = 0.229\; \hat{I}\, \hat{E} \qquad (10r)$$

BASIC RECTIFIER POWER CONVERTERS

The conversion of ac power to dc power can be classified into three general types of circuits. They are identifiable by the output circuit as seen by the rectifier tube. The output may be either an unfiltered resistance load, Fig. 8A, a choke-input-filtered load, Fig. 8B, or a capacitor-input-filtered load, Fig. 8C. Also shown with these three typical circuits are their significant voltage-and-current-vs.-time characteristics. For simplicity, it has been assumed that the rectifying diodes are ideal switches being "on" only when the ac supply voltage E is more positive than the voltage e_o at the rectifier output. It was also assumed that the choke and capacitor are very large.

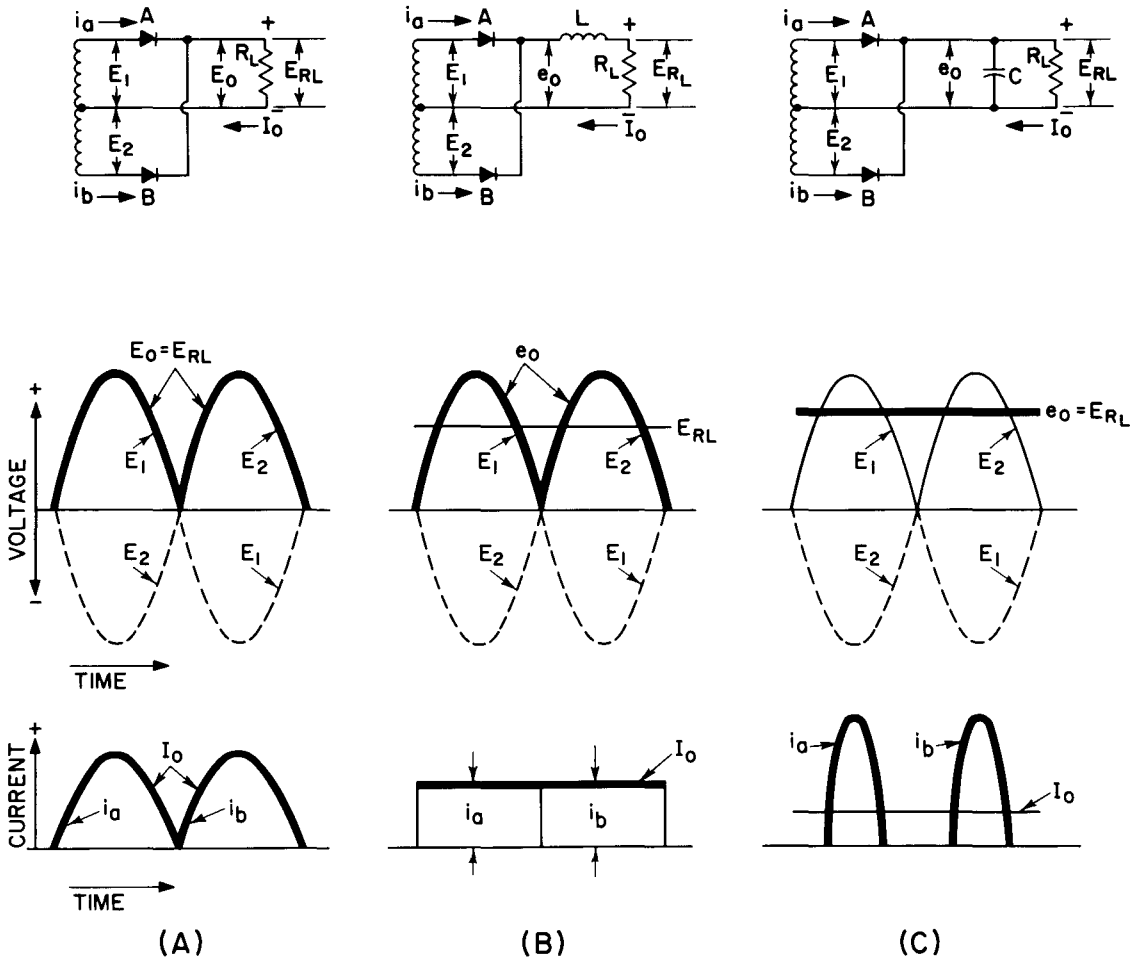


Figure 8. Comparison of Voltages and Currents in a Single-Phase Full-Wave Rectifier; (A) no filter, (B) choke-input filter, (C) capacitor-input filter

In steady-state operation, the following differences may be noted in these three types of circuits: (1) no filter, resistance load only, (2) choke-input-filtered load, and (3) capacitor-input filtered load.

No Filter, Resistance Load Only. Both the output voltage e_o and the output current I_o are unidirectional pulsating half-sinusoids at the rectifier output terminals, Fig. 8A. In multiple-phase operation (see Figs. 12, 13, 14) the overlapping of the phases reduces the pulse-amplitude of both e_o and I_o.

Choke-Input-Filtered Load. The output current I_o is constant, but the output voltage e_o is a series of unidirectional half-sinusoids at the rectifier terminals, Fig. 8B. Each diode passes a rectangular-shaped current pulse i_a and i_b, respectively. The voltage across the load resistor R_L is constant, because all of the ac voltage components of e_o appear across the large choke L.

Capacitor-Input-Filtered Load. The output voltage e_o has a constant amplitude, but the output current i_o consists of a series of unidirectional pulsating fractional half-sinusoids at the rectifier terminals, Fig. 8C. The current I_o through the load resistor R_L is constant, because all the ac components of the rectifier output current are bypassed through the large capacitor C. Note that, for the assumed ideal rectifier switches, the dc output voltage e_o should just equal the peak of the input ac voltage E_1 and E_2. Actually, the dc output voltage e_o is very sensitive to even small amounts of resistance in series with the diodes. In Fig. 8C, therefore, a small resistor is tacitly assumed to be in series with each diode.

RESISTANCE LOAD AND CHOKE-INPUT-FILTERED LOAD

The operation of resistance load, and choke-input-filtered load circuits as applied to rectifiers has been understood for many years.[3, 4, 5] Both types of loading have identical waveshapes at the rectifier output terminals. These waveshapes are indicated as e_o in Figs. 9 through 14; they are either a series of repeating half-sinusoids, as in Figs. 9, 10, 11, or, the top parts of repeating half-sinusoids as in Figs. 12, 13, 14. By application of the Fourier analysis, these waveshapes may be resolved into a dc component and a series of ripple voltages which are harmonics of the supply frequency, as follows

For single-phase, half-wave connections (Fig. 9)

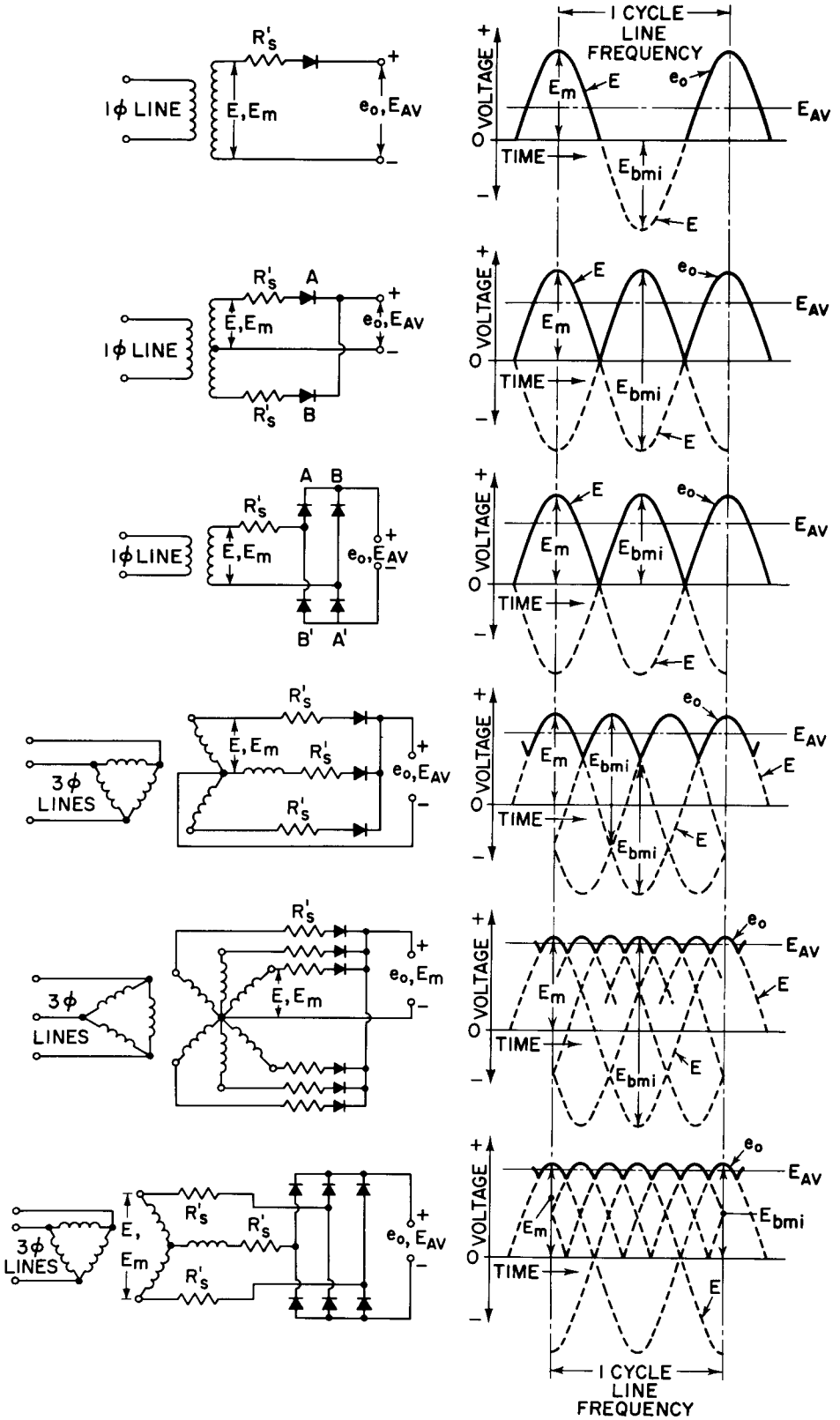


Figure 9. Single-Phase, Half-Wave Rectifier (not used with choke-input filter)

Figure 10. Single-Phase, Full-Wave Rectifier

Figure 11. Single-Phase, Bridge Rectifier

Figure 12. Three-Phase, Half-Wave Rectifier

Figure 13. Three-Phase, Full-Wave Rectifier

Figure 14. Three-Phase Bridge Rectifier

In Figs. 9 to 14 the voltage curves hold for both resistance load and choke-input-filtered load. The heavy solid curve (e_o) is the voltage at the rectifier output terminals. The voltage consists of a dc component E_{Av} (dot-dash line) and a superimposed ripple voltage. E and E_m are the RMS and peak secondary voltages, respectively. E_{bmi} is the peak inverse voltage across a diode.

Table I

Normalized Characteristics for Rectifier Circuits

With Resistance and Choke-Input-Filtered Loads▲

E = Transformer Secondary Voltage (rms)	I_{av} = Average DC Output Current
E_{av} = Average DC Output Voltage	I_b = Average Plate Current
E_m = Peak Transformer Secondary Voltage	I_p = Plate Current (rms)
E_{bmi} = Peak Inverse Plate Voltage	I_{pm} = Peak Plate Current
E_r = Major Ripple Voltage (rms)	P_{ap} = Transformer Primary Volt-Amperes
f = Supply Frequency	P_{as} = Transformer Secondary Volt-Amperes
f_r = Major Ripple Frequency	P_{dc} = DC Power = (E_{av} x I_{av})

Item	1-Phase Half-Wave (Fig. 9)	1-Phase Full-Wave (Fig. 10)	1-Phase Bridge (Fig. 11)	3-Phase Half-Wave (Fig. 12)	3-Phase Full-Wave (Fig. 13)	3-Phase Bridge (Fig. 14)
Voltage Ratios						
E_{av}/E_m	0.318	0.636	0.636	0.827	0.955	0.955
E_{av}/E	0.451	0.900	0.900	1.130	1.350	1.350
E_{bmi}/E	1.41	2.83	1.41	2.45	2.83	1.41
E_{bmi}/E_{av}	3.14	3.14	1.57	2.09	2.09	1.05
E_r/E_{av}	1.11	0.471	0.471	0.177	0.040	0.040
Frequency Ratio						
■ f_r/f	1	2	2	3	6	6
Current Ratios						
■ I_b/I_{av}	1	0.5	0.5	0.333	0.167	0.333
Resistive Load						
I_p/I_{av}	1.57	0.785	0.785	0.587	0.409	0.578
I_{pm}/I_{av}	3.14	1.57	1.57	1.21	1.05	1.05
I_{pm}/I_b	3.14	3.14	3.14	3.63	6.3	3.14
Inductive Load○						
I_p/I_{av}	*	0.707	0.707	0.577	0.408	0.577
$I_{pm}I_{av}$	*	1.00	1.00	1.00	1.00	1.00
Power Ratios:						
Resistive Load						
P_{as}/P_{dc}	3.49	1.74	1.23	1.51	1.82	1.063
P_{ap}/P_{dc}	2.69	1.23	1.23	1.24	1.28	1.063
Inductive Load○						
P_{as}/P_{dc}	*	1.57	1.11	1.48	1.28	1.05
P_{ap}/P_{dc}	*	1.11	1.11	1.21	1.05	1.05

Notes:

▲ Conditions assume sine-wave voltage supply; zero voltage drop across rectifiers when conducting; no losses in transformer or choke (R_S' = 0 in Figs. 9 to 14); output load is a pure resistance.

○ The use of a large filter-input choke is assumed.

* Single-phase, half-wave, choke-input-filtered load has no practical significance; only a minute pulsating dc current will flow.

■ These ratios also apply for the case of capacitor-input filtered load.

$$e_o = 0.451E(1 + 1.571 \cos \omega t + 0.667 \cos 2\omega t - 0.133 \cos 4\omega t + \ldots) \quad (11)$$

For both full-wave or bridge single-phase connections (Figs. 10, 11)

$$e_o = 0.900E(1 + 0.667 \cos 2\omega t - 0.133 \cos 4\omega t + \ldots) \quad (12)$$

For three-phase, half-wave connections (Fig. 12)

$$e_o = 1.130E(1 + 0.250 \cos 3\omega t - 0.057 \cos 6\omega t + \ldots) \quad (13)$$

For both full-wave or bridge three-phase connections (Figs. 13, 14)

$$e_o = 1.350E(1 + 0.057 \cos 6\omega t - 0.014 \cos 12\omega t + \ldots) \quad (14)$$

where

e_o = voltage at rectifier output terminals

E = rms voltage per plate at the input to rectifier

ω = angular frequency ($2\pi f$) at the applied line frequency f

t = time, in seconds.

The first term on the right hand side of these equations gives the dc voltage available as a result of the rectification process; the remaining series of terms expressed as cosine functions represent the usually undesirable ripple voltage.

Because the output voltage e_o at the rectifier can be expressed analytically, as indicated by Eqs. (11), (12), (13), and (14), it is possible and convenient to replace the whole rectifier circuit by equivalent circuits (see Figs. 15, 16, and 17). In the equivalent circuits the output loading of the original circuit remains unchanged but the rectifiers and ac supply voltage are replaced by a battery E_{av}, a series resistor R_s, and a string of ripple voltage generators E_1, E_2, E_3, etc., all of which are in series with the output loading. The effective series resistor R_s consists of two parts, r_s and r_d. The first part r_s is the sum of all the linear resistances in series with each diode but does not include the load resistor R_L; an equivalent transformer resistance, the dc resistance of filter chokes, and any protective surge resistances would be included. The second part r_d is the equivalent diode resistance and is nonlinear.

The operation characteristic for the choke-input filter circuits of Figs. 16 and 17 is the relationship between the dc output current I_o and the dc output voltage E_o for a fixed ac supply voltage. For its determination, the ac ripple voltage generators may be disregarded. Because the diode, while in the conduction state, must pass the total output current I_o, the expression for the dc output voltage is given by

$$E_o = E_{av} - I_o r_s - e_d \quad (15)$$

where E_{av} is the average dc voltage output indicated in Table I, as well as the first right-hand term of the applicable Fourier series (Eqs. 11, 12, 13, or 14); r_s is

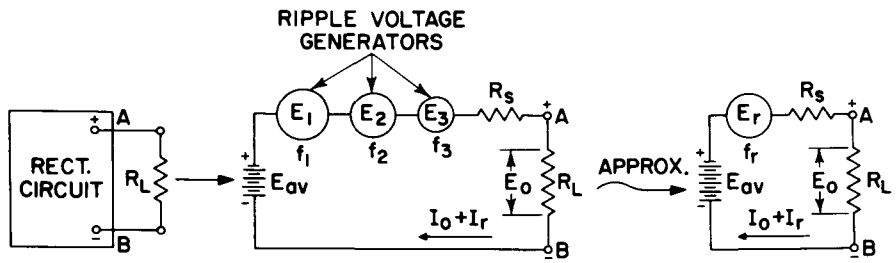


Figure 15. Equivalent Circuit for Rectifier with Resistance Load

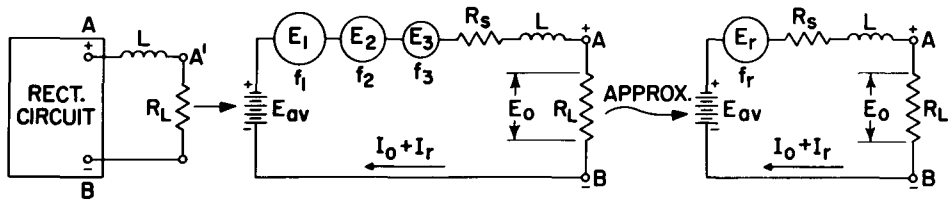


Figure 16. Equivalent Circuit for Rectifier with Choke-Input-Filtered Load

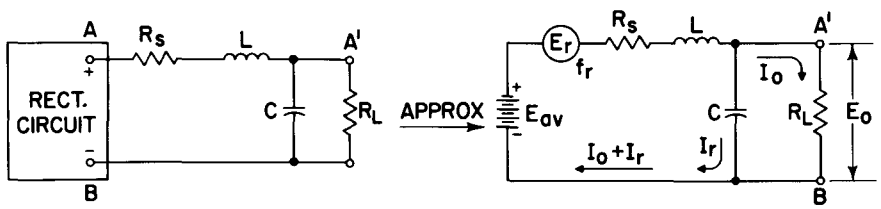


Figure 17. Equivalent Circuit for Rectifier with Choke-Input-Filtered Load and a Capacitor C Added to Reduce Ripple

the linear resistance in series with each diode, and e_d is the diode voltage drop corresponding to the full load current I_o. The graphical determination of this operation characteristic is shown in Fig. 18, where e_{R_S} is the voltage drop across the diode and any added series resistance r_s, e_d is the voltage drop across the diode, and e_{r_s} is the voltage drop across r_s. The slope of R_L is the load resistance and its intersection with the output curve at E_o determines the dc output current I_o. Fig. 19 shows the operation characteristics corresponding to Fig. 17.

The peak current I_m which flows through a single diode in choke-input-filtered operation cannot exceed twice the total output current I_o (see Fig. 20), since this would require the diodes passing current in their non-conducting direction. Its actual magnitude is determined from the peak ripple current $\hat{I}_r$ superimposed on the dc output current I_o,

$$I_m = I_o + \hat{I}_r \tag{16}$$

The peak ripple current can be determined from the equivalent circuit diagrams of Figs. 16 or 17. If the practical assumptions are made that $R_S \ll R_L$ and that the capacitor impedance Z_c, at the main ripple frequency f_r, is less than Z_L and R_L, then, for the choke circuit in Fig. 16

$$\frac{\hat{I}}{I_o} = \sqrt{2}\,\frac{R_L}{\sqrt{Z_L^2 + R_L^2}} \cdot \frac{E_r}{E_{av}} \tag{17}$$

and for the choke-capacitor circuit in Fig. 17

$$\frac{\hat{I}}{I_o} = \sqrt{2}\,\frac{R_L}{Z_L} \cdot \frac{E_r}{R_{av}} \tag{18}$$

where Z_L = choke impedance $2\pi f_r L$ and f_r is the main ripple frequency

E_r/E_{av} = ratio of main ripple rms voltage to the dc-output voltage (see Table I)

$\hat{I}/I_o$ = ratio of peak ripple current to dc output current.

When the choke alone is used for filtering the output (see Fig. 16), the ratio of peak ripple to dc output current can never exceed the value 0.667 in both full-wave and bridge single-phase circuits. However, this condition is not the case when a large capacitor C is shunted across R_L to further reduce ripple voltages at the output terminals. Inspection of the equivalent circuit in Fig. 17 will indicate that, with the assumptions made above, the magnitude of the ripple current is limited solely by the choke inductance L. If the computed ratio $\hat{I}/I_o$ is greater than unity, then the impossible condition occurs as indicated by curve 4 in Fig. 20. The rectifiers cannot pass current opposite to the normal conduction, as demanded by the dotted part of this curve. For a given load resistor, there is a correlated minimum inductance which just allows the peak ripple current to equal the dc output current (see curve 3, Fig. 20). Its value is indicated for the choke and capacitor circuit (Fig. 17) by

$$L_{min} = \frac{R_L}{n\pi f_1} \text{ henries} \tag{19}$$

where n = 6 for both single-phase full-wave and bridge circuits (Figs. 10 and 11)

n = 24 for three-phase half-wave circuit (Fig. 12)

n = 210 for both three-phase full-wave and bridge circuits (Figs. 13 and 14)

and R_L and f_1 are the load resistor and the line frequency, respectively.

Note, that as the dc output current is decreased (R_L is increased), this minimum inductance value must increase. In Fig. 19, point A corresponds to operation with L_{min}. In the region A-B, operation is under normal choke-input filter conditions. In the region 0-A, operation is similar to that of a capacitor-input-filtered load circuit. For a practical quantitative analysis of the operation with choke inductance less than the minimum critical value, refer to Heymann.[6]

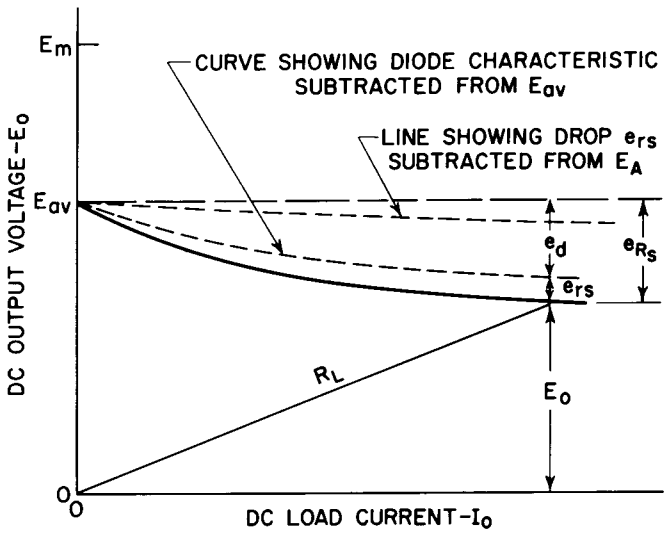


Figure 18. Construction of Operation Characteristic for Resistance Load (Fig. 15) and for Choke-Input-Filtered Load (Fig. 16)

The power dissipation in the diodes and in any series resistors of the choke-input-filter circuits is determinable by first assuming that only dc current flows through the choke. A correction factor is then applied which increases this dissipation to account for the effect of adding a ripple component to the dc current; the following expressions may then be obtained:

$$W_{pp} = \frac{k}{N} I_o e_d \text{ watts per plate, for diodes} \tag{20}$$

$$\text{and } W_{R_S} = \frac{k}{M} I_o^2 R_S \text{ watts for linear series resistance} \tag{21}$$

where k is the correction factor due to ripple current (it may be read from Fig. 21), I_o is the total dc output current, and e_d is the voltage drop across the diode for total current I_o and

N = 2, for both full-wave and bridge single-phase circuits (Figs. 10, 11)

N = 3, for both half-wave and bridge three-phase circuits (Figs. 12, 14)
N = 6, for full-wave three-phase circuits (Fig. 13).

For a series resistor carrying full-load current full time (i. e., the series dc resistance of the choke), M is equal to 1.

For the series resistors marked R_S' in Figs. 10 through 14:

M = 1, for single-phase bridge circuit (Fig. 11)
M = 2, for single-phase full-wave circuit (Fig. 10)
M = 3, for both half-wave and bridge three-phase circuits (Figs. 12, 14)
M = 6, for three-phase full-wave circuit (Fig. 13)

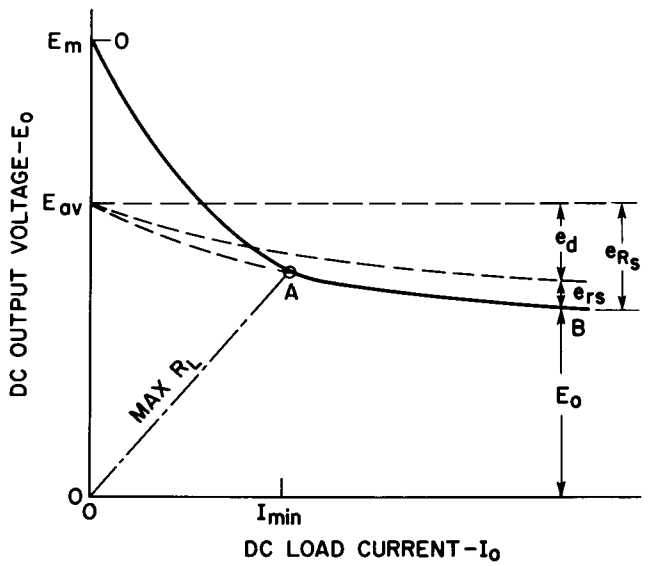


Figure 19. Operation Characteristic for Choke-Input-Filtered Load with Large Capacitor C added Across the Load Resistor (Fig. 17)

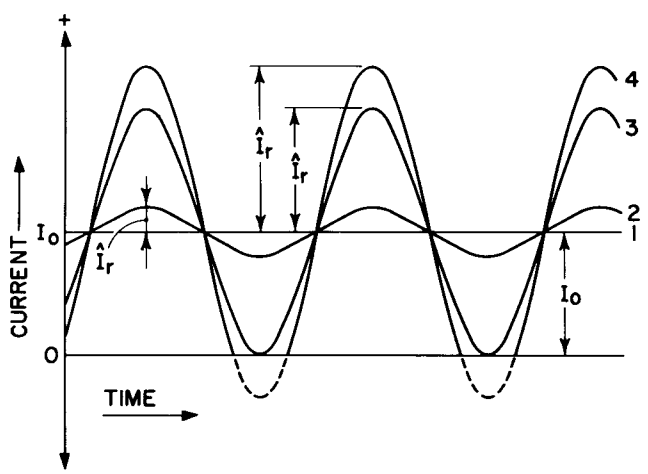


Figure 20. Current Waveshapes Flowing in Choke, Choke-Input Filter Circuit with Capacitor Added (Fig. 17); Showing Effect of Varying the Choke Inductance L

A peak inverse voltage E_{bmi} appears across each rectifying element one or more times during a cycle of line frequency. Its voltage polarity is opposite to that for diode conduction, that is, the plate will be highly negative with respect to the cathode. It is important that the rectifier diodes withstand this peak inverse voltage, otherwise, should breakdown occur, a resulting short-circuit will appear across the ac supply lines. For example, the peak inverse voltage in the single-phase full-wave circuit in Fig. 10 may be determined by considering the rectifiers A and B as switches that are closed only during their conduction time interval. If it is assumed that switch A is closed, then the maximum voltage across the open switch B will be twice the peak voltage of the transformer's half secondary winding ($2\sqrt{2}E$). Bridge circuits show to advantage by having a peak inverse voltage only half that of the full-wave circuits. For an example of this benefit, refer to the single-phase bridge circuit in Fig. 11. Note that when switches A and A' are closed, the open switches B and B' will each have a peak inverse voltage equal to the peak value of the ac supply line $\sqrt{2}E$ across them. The peak inverse voltages are indicated as E_{bmi} in the waveshape diagrams of Figs. 9 through 14 and are normalized in Table I.

TUNGSTEN-LAMP-LOADED RECTIFIER

This case is very special and of particular interest in tube manufacturing, because tungsten lamps are used to stabilize the tube characteristics during the aging of diodes.

Barbrow and Meyer[7] give wattage, current, and voltage data for tungsten-filament vacuum and gas-filled lamps in the range of 15 to 500 watts. This information and that of Langmuir[8] and Haller[9] produce a single-valued curve when normalized; see the middle curve of Fig. 22. For this curve, the rms current and rms voltage of the lamp have been normalized to the rated nominal operating current and voltage, respectively, and drawn in graph form.

When a tungsten lamp is used as the load in either a half-wave or full-wave rectifier, the dissipation and currents of the diode can be determined from two voltage readings. One reading is the rms voltage across the lamp and the other is the rms voltage of the supply. (Do not use a rectifier-type ac meter or a Volt-ohmyst to make the reading across the lamp because they are calibrated to measure only pure sine or cosine voltage waveshapes.) The plate dissipation and diode currents can be determined from two measured readings with the aid of Figs. 22 and 23.

The curves in Fig. 23 were computed from

$$W_d/W_n = \frac{I_{rms}}{I_N}\left[\frac{E_{ac}}{E_N} - \frac{E_L}{E_N}\right] \quad \text{for full-wave operation,} \tag{22}$$

and

$$W_d/W_n = \left[\frac{I_{rms}}{I_N}\left[\frac{E_{ac}}{\sqrt{2}\,E_N} - \frac{E_L}{E_N}\right]\right] \quad \text{for half-wave operation} \tag{23}$$

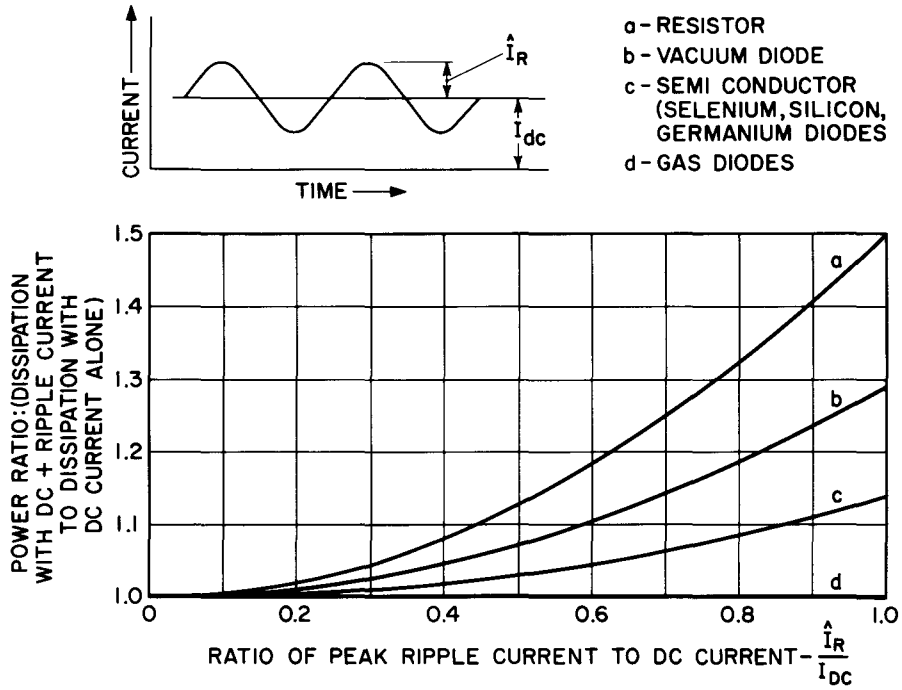


Figure 21. Increase in Dissipation Due to Presence of Ripple Current Superimposed on a DC Current

In these equations the ratio I_{rms}/I_N is a function of E_L/E_N; correlation of the two ratios is shown by the middle curve of Fig. 22.

CAPACITOR-INPUT-FILTERED LOAD

In making a rigorous analytical solution of the rectifier circuit with a capacitor-input-filtered load, many difficulties will be encountered when transcendental, trigonometric, and exponential functions appear together. Some of the methods of obtaining a practical solution to this type circuit are indicated in the references.[5, 6, 10, 11, 12, 13] Particular note should be made of the Schade paper[5] in which no restrictions are placed on the circuit parameters.

Fig. 24 shows the currents and voltages occurring in a half-wave single-phase capacitor-input load circuit, which operates in two periods. These periods are identifiable as a capacitor-changing interval (1a to 2a, 1b to 2b, 1c to 2c) and a capacitor-discharging interval (2a to 3a, 2b to 3b, 2c to 3c). The diode rectifier D, acting as a switch, closes when the ac supply voltage E rises above the capacitor voltage e_0; it opens again when E drops below the capacitor voltage. The output waveshape e_0 is a function of the filter-capacitor charging and discharging time constants, R_SC and R_LC, respectively. To demonstrate the effect of changing these time constants in the circuit of Fig. 24, all the parameters are held constant except the value of the filter capacitance C. Then

for condition a, the discharge time constant $R_LC = 300\ T_r$
for condition b, the discharge time constant $R_LC = 3T_r$
for condition c, the discharge time constant $R_LC = 0.3\ T_r$
where T_r is one period of the ripple frequency.

The load resistor R_L is approximately equal to $10R_S$ and therefore the charging time constants are approximately one-tenth those of the corresponding discharge time constants. The diode D is assumed to be an ideal switch. When both time constants are very long compared to a cycle of line frequency (as in condition a), the voltage of the filter capacitor e_0 remains substantially constant during the charge and discharge periods and, consequently, there is negligible ripple voltage. With both time constants short compared to a cycle of line frequency (as in condition c), the voltage across the filter capacitor e_0 is far from constant, dropping almost to zero volts at the end of the discharge period (points 1c, 3c). Because the charging time constant is small ($0.03T_r$), the filter capacitor builds up its voltage rapidly (period 1c to 2c). The average dc capacitor voltage E_0 is lower than for condition a, and has a large superimposed ripple voltage superimposed upon it. Condition b, with its time constants for charging and discharging the filter capacitor intermediate between those of conditions a and c, represents a practical condition. Note that the dc output voltage and the peak and average currents are the same as for condition a. The two conditions differ by a small phase shift in the charging currents, and some not negligible ripple voltage present in condition b.

The magnitude of the average output voltage E_0 in each condition must adjust itself to a level such that the average value of the pulse charging currents just equals their respective dc output current I_0. For the cases illustrated in Fig. 24, the area under a single pulse of

rectifier current i_d must equal the area under its associated dc output current (line I_O), for one repeating period T_r. Because the pulse of charging current flows only through the series resistor R_S and the dc output current flows only through the load resistor R_L, the dc output voltage E_O will be critical to both these resistance values. Figs, 25, 26, 27 show how sensitive the output voltages are to parameter $\bar{R}_S/R_L$.

When Schade[5] investigated the operation of capacitor-

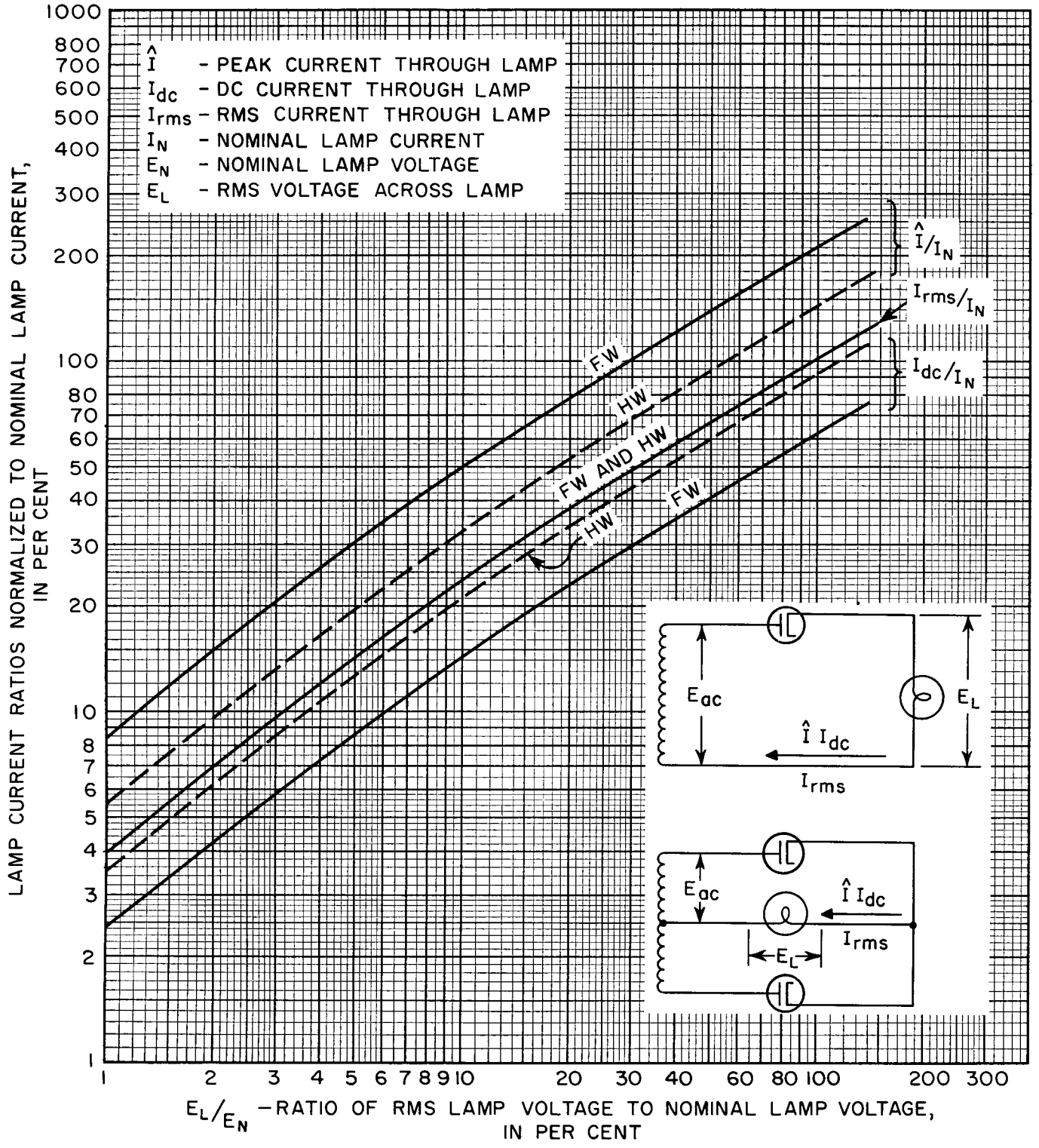


Figure 22. Relation of Current and Voltage in a Single Tungsten Lamp Used as Load in Half-Wave and Full-Wave Rectifiers

filter rectifier circuits (in the middle 1930's), he found he could replace the diode under investigation with an "ideal switch" and a linear equivalent resistance by adding diodes in parallel with the original diode until the observed peak, rms, and dc currents each became constant. This constancy was always at a level higher than the currents of the single diode. By the addition of series resistance to the diode, these currents could then be brought back to their original value. But a different value of resistance was needed to bring the peak, rms, or dc currents, respectively, back to normal. The relationship between these equivalent resistances was resolved by careful analysis of the many observed current waveshapes and gave the following important correlations:

$$\hat{r}_d = 0.88\,\bar{r} = 0.935\,|r| = \frac{\hat{e}_d}{\hat{i}_p} \qquad (24)$$

where $\hat{e}_d$ = peak diode voltage drop (corrected for contact potential, if necessary) for peak current $\hat{i}_p$

$\hat{r}_d$ = peak diode resistance

$\bar{r}_d$ = equivalent average diode resistance

$|r_d|$ = equivalent rms diode resistance

Note: If an additional resistance R_s is present in series with the diode, then the above resistances must be replaced by

$$\hat{R} = R_s + \hat{r}_d$$

$$\bar{R} = R_s + \bar{r}_d$$

$$|R| = R_s + |r_d|$$

The operational characteristics of Figs. 25 through 29 were constructed by Schade[5] using these equivalent re-

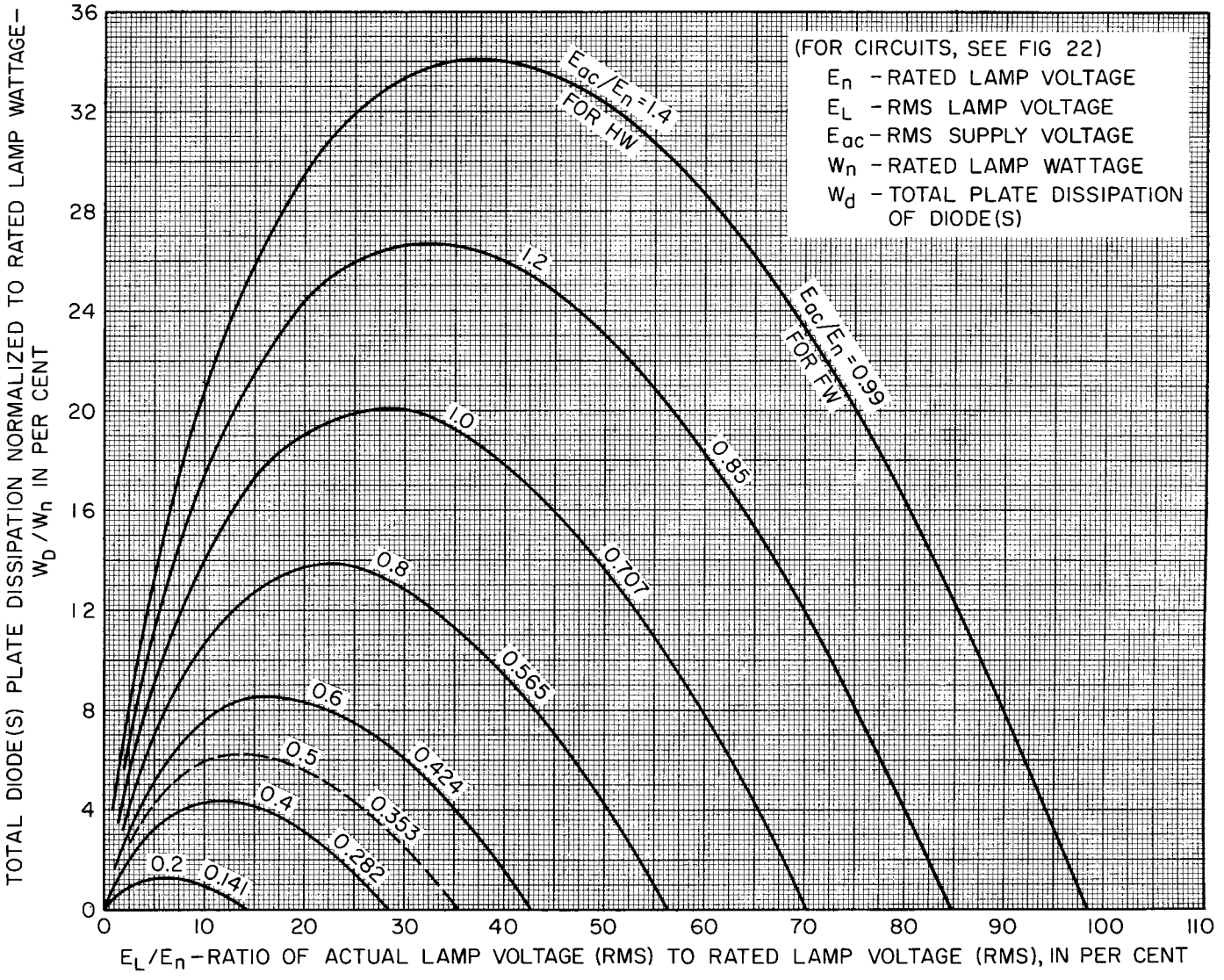


Figure 23. Total Anode Dissipation in Half-Wave and Full-Wave Rectifier Circuits with a Tungsten Lamp Load

sistance concepts. These characteristics are unique in that no qualifying restrictions are placed on any of the circuit parameters. The only assumption made is that the supply voltage must have a sinusoidal waveshape.

CAPACITOR-INPUT-FILTER CIRCUITS USING LARGE VALUES OF CAPACITANCE

If the filter capacitance is large (so that the output voltage is pure dc) the capacitor-input-filter circuit may be handled analytically. Even the presence of some ripple voltage does not alter the values of the dc output voltage, the dc output current, and the peak current, or its current and voltage waveshapes, as can be seen by comparing conditions a and b in Fig. 24. Condition b is identical to condition a except for the presence of a ripple voltage and a small phase shift in the position of the diode current pulses. Voltage conditions can then be simplified as shown in Fig. 30. The family of operation characteristics in Figs. 32 through 38 was

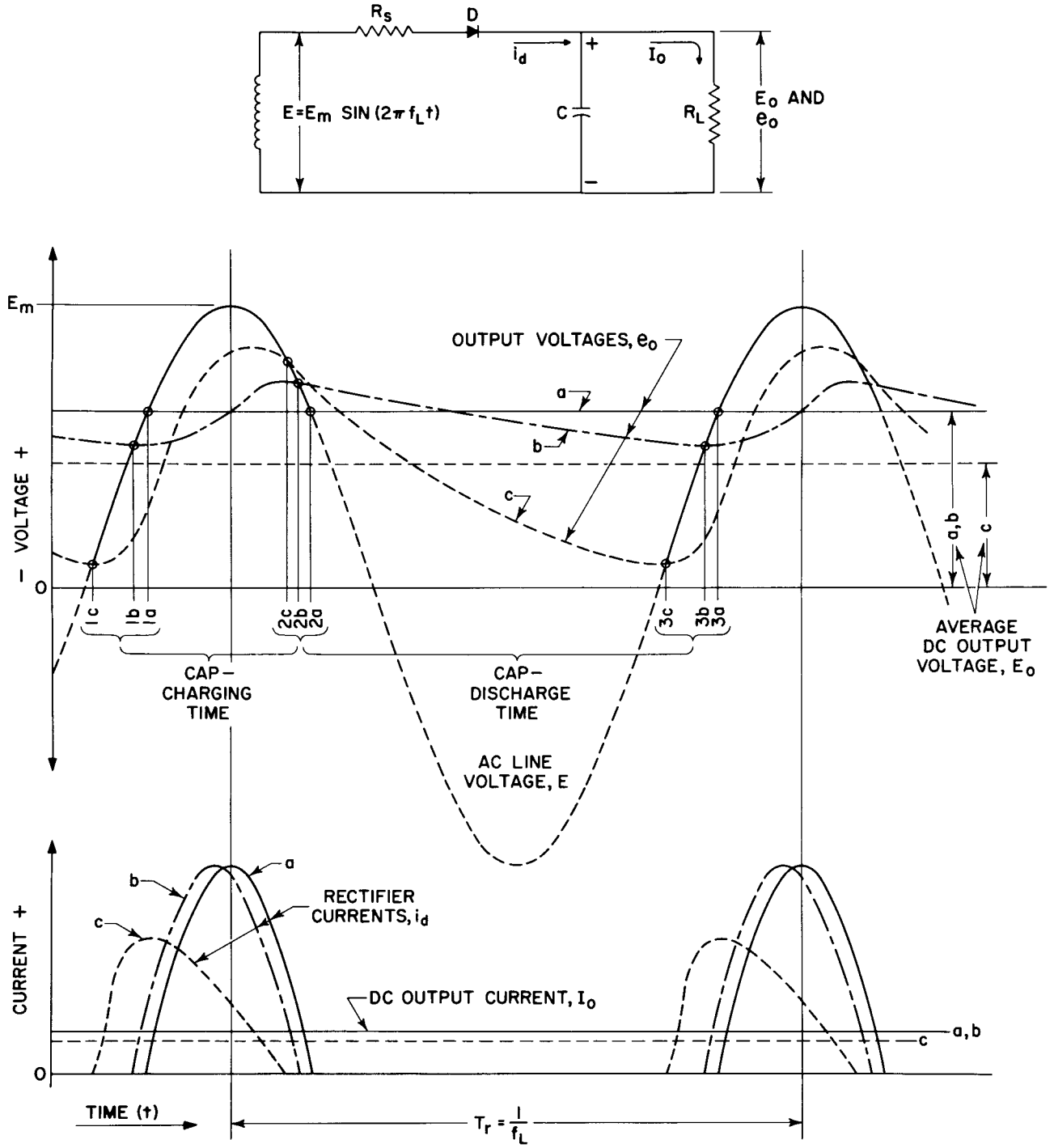


Figure 24. Effect on Current and Voltage Waveshapes of Changing the Size of the Input-Filter Capacitor in a Half-Wave, Single-Phase Rectifier

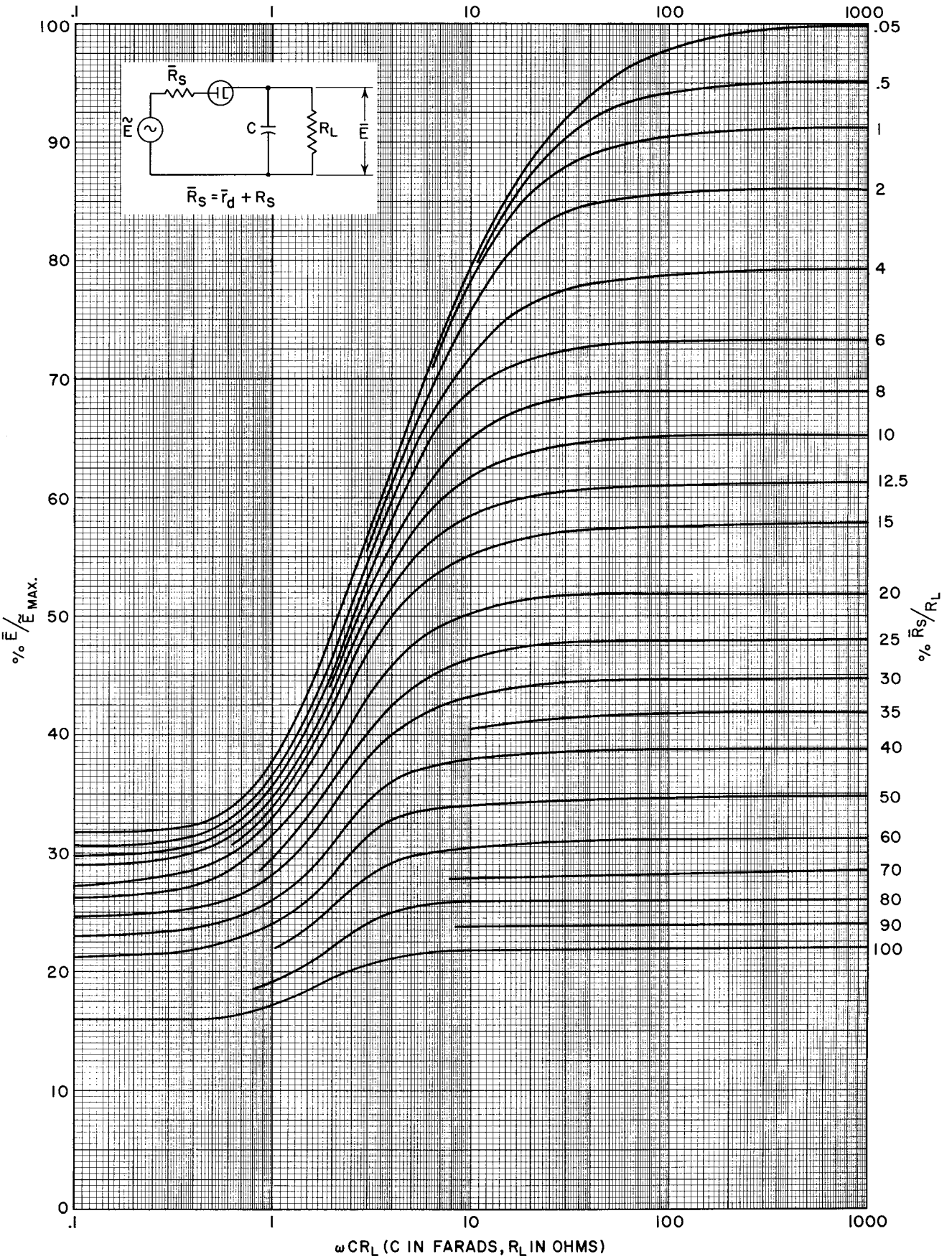


Figure 25. Relation of Applied AC Voltage to DC-Output Voltage in Half-Wave Rectifier

developed with the aid of Eqs. (9) and (10) and by considerable manipulation. To insure that the filter capacitance actually is sufficiently large, consult Fig. 31. With the capacitance at the minimum value computed from this curve, the dc output voltage will be one per cent less than that for a capacitance of infinite value. The corresponding ripple voltage present is also indicated. Note that in Fig. 31 the ordinate $R_L C f_r$ is a dimensionless numeral. It is the ratio of the discharge time constant to the period of a charge-discharge cycle. The ripple voltage curve in Fig. 31 indicates the tolerable degree of deviation from the ideal assumption of pure dc output within the characteristics curves of Figs. 37 and 38.

A step-by-step process is shown here for computing an operating point for the large capacitor-filter circuits. Inspection of Fig. 30 indicates that the dc output voltage E_o can be expressed by

$$E_o = E_m - (\hat{e}_d + \hat{e}_r) = E_m - \hat{e}_d - \hat{I}R_s \qquad (25)$$

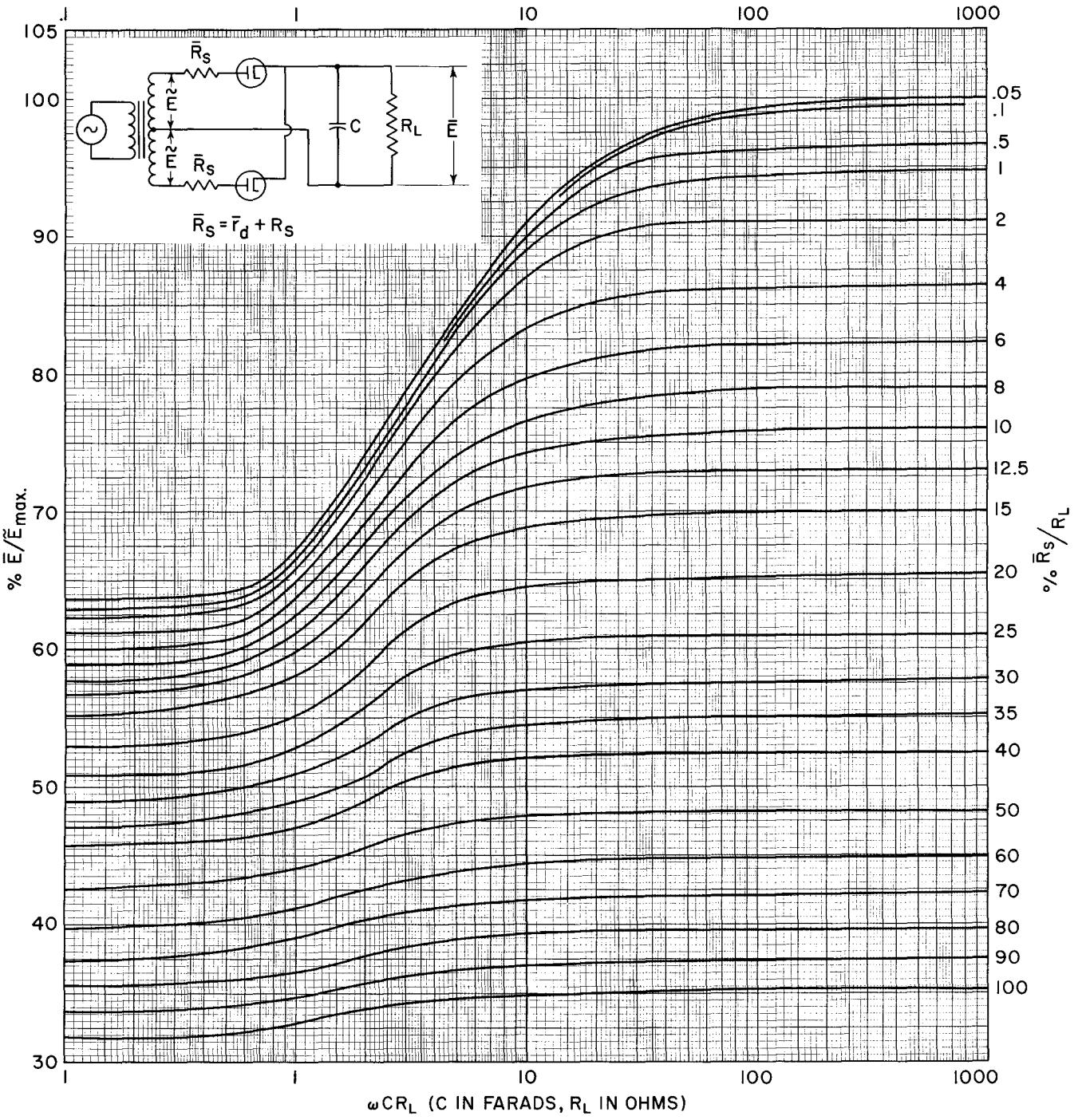


Figure 26. Relation of Applied AC Voltage to DC-Output Voltage in Full-Wave Rectifier

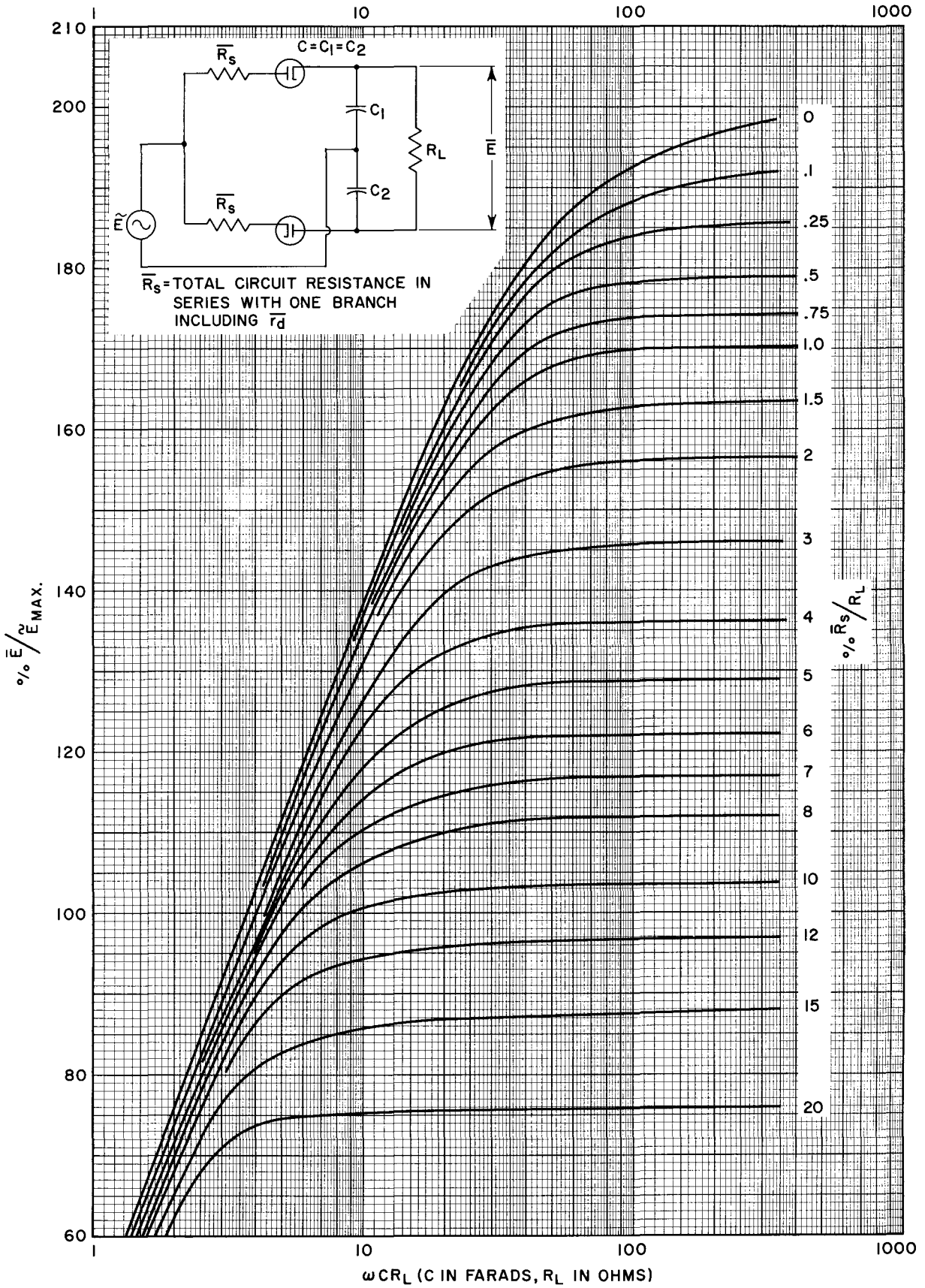


Figure 27. Relation of Applied AC Voltage to DC-Output Voltage in Voltage-Doubling Rectifiers

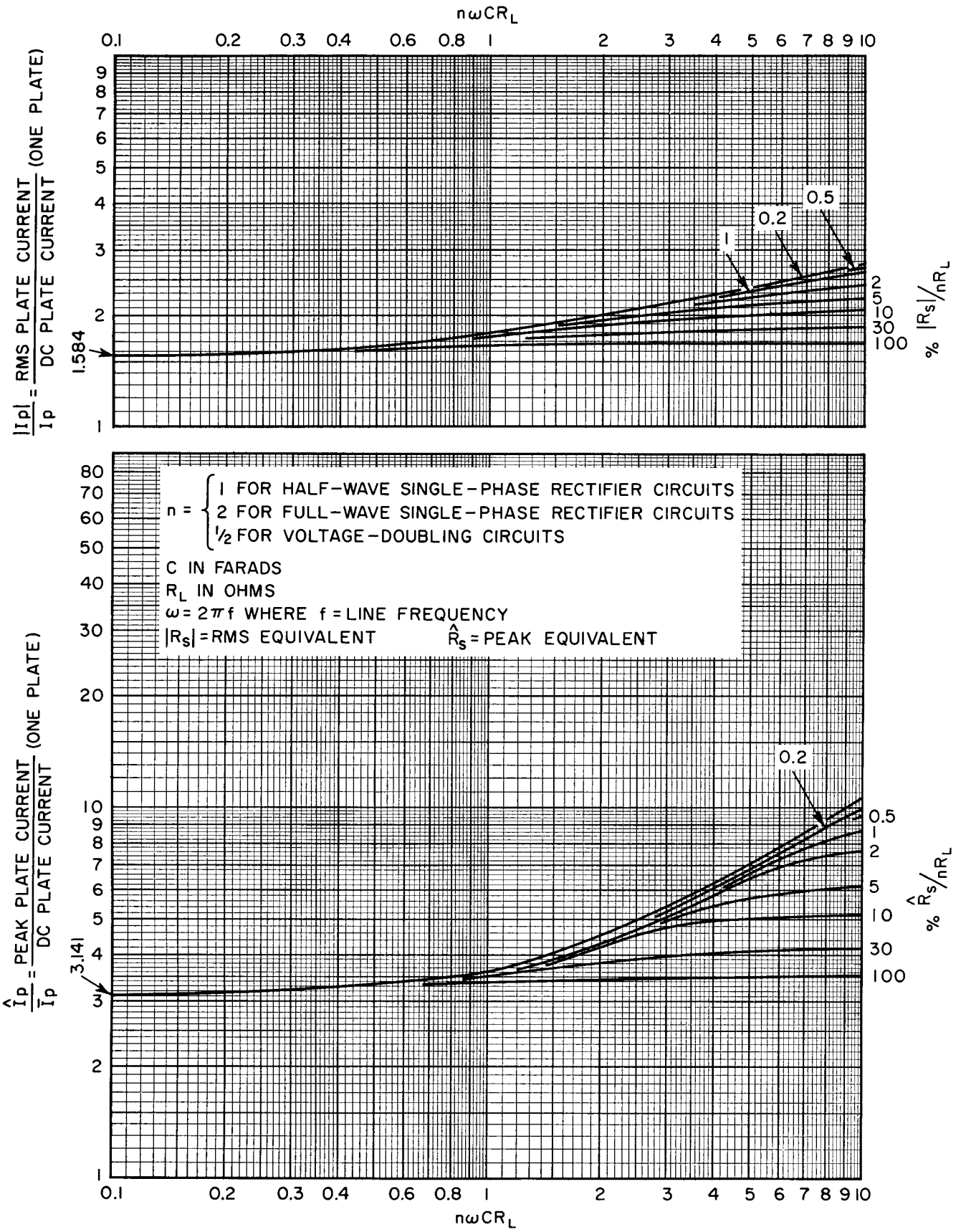


Figure 28. Relations of Peak, RMS, and Average Diode Currents

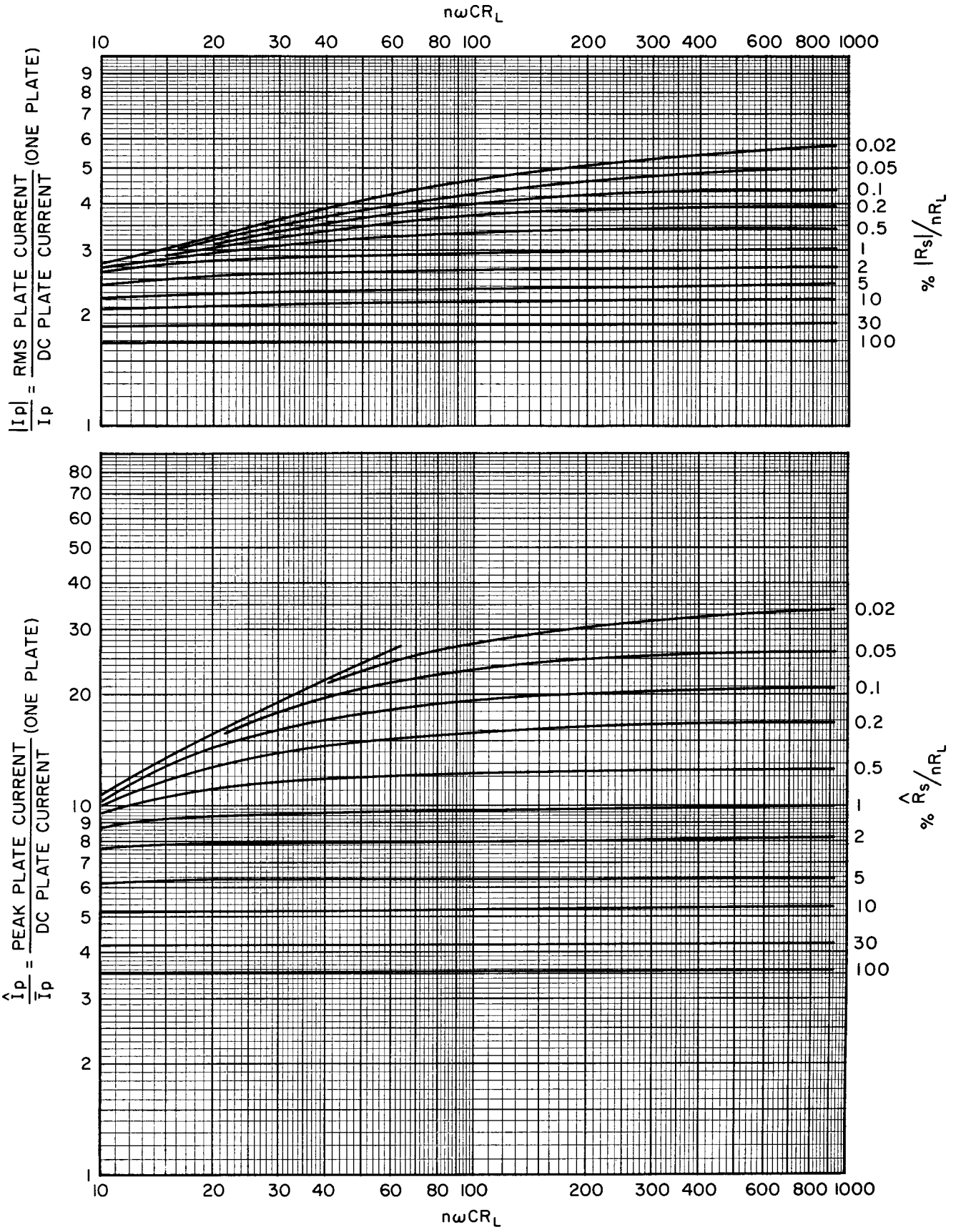


Figure 28. Relations of Peak, RMS, and Average Diode Currents

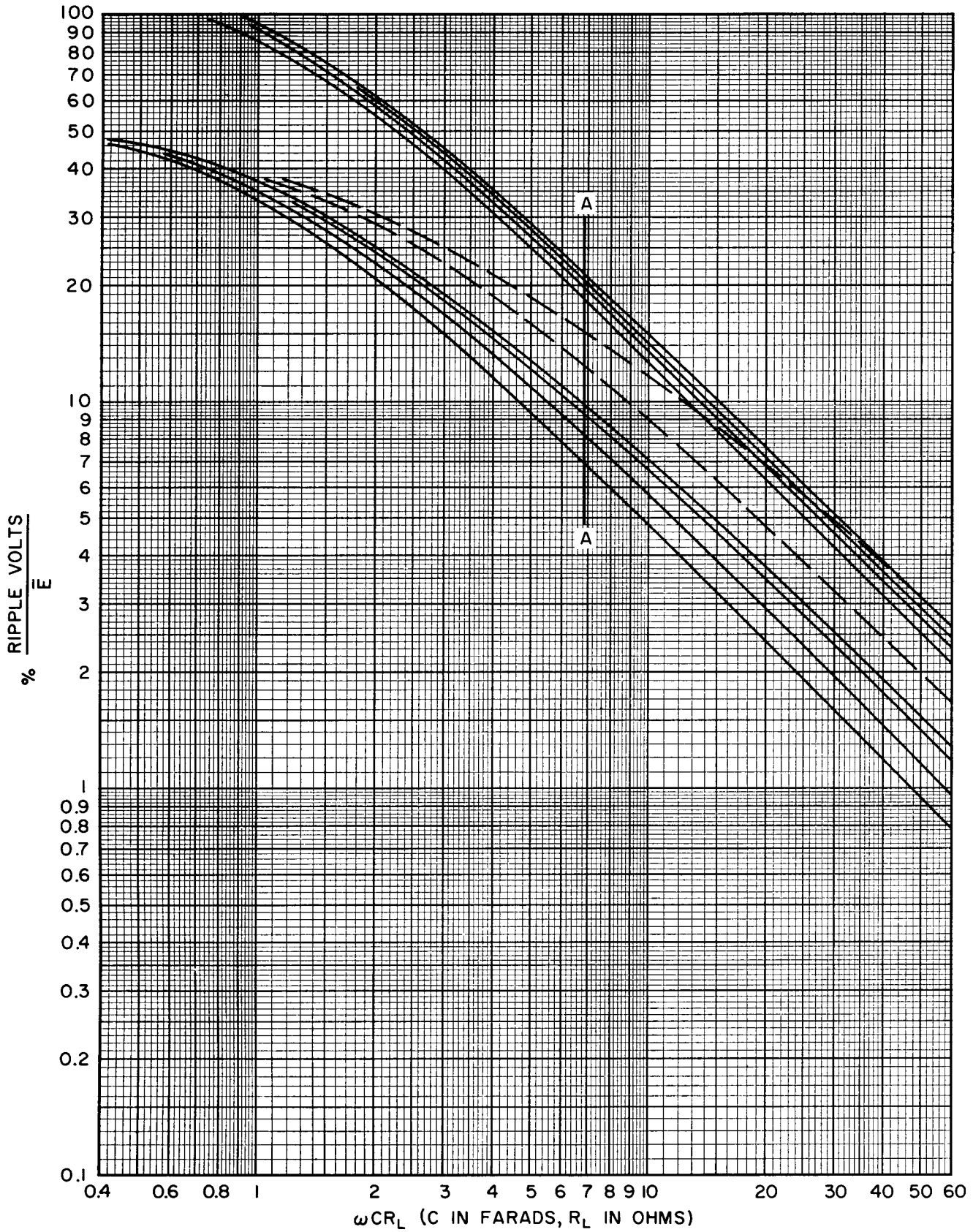


Figure 29. RMS Ripple Voltage for Capacitor-Input-Filtered Load Circuits

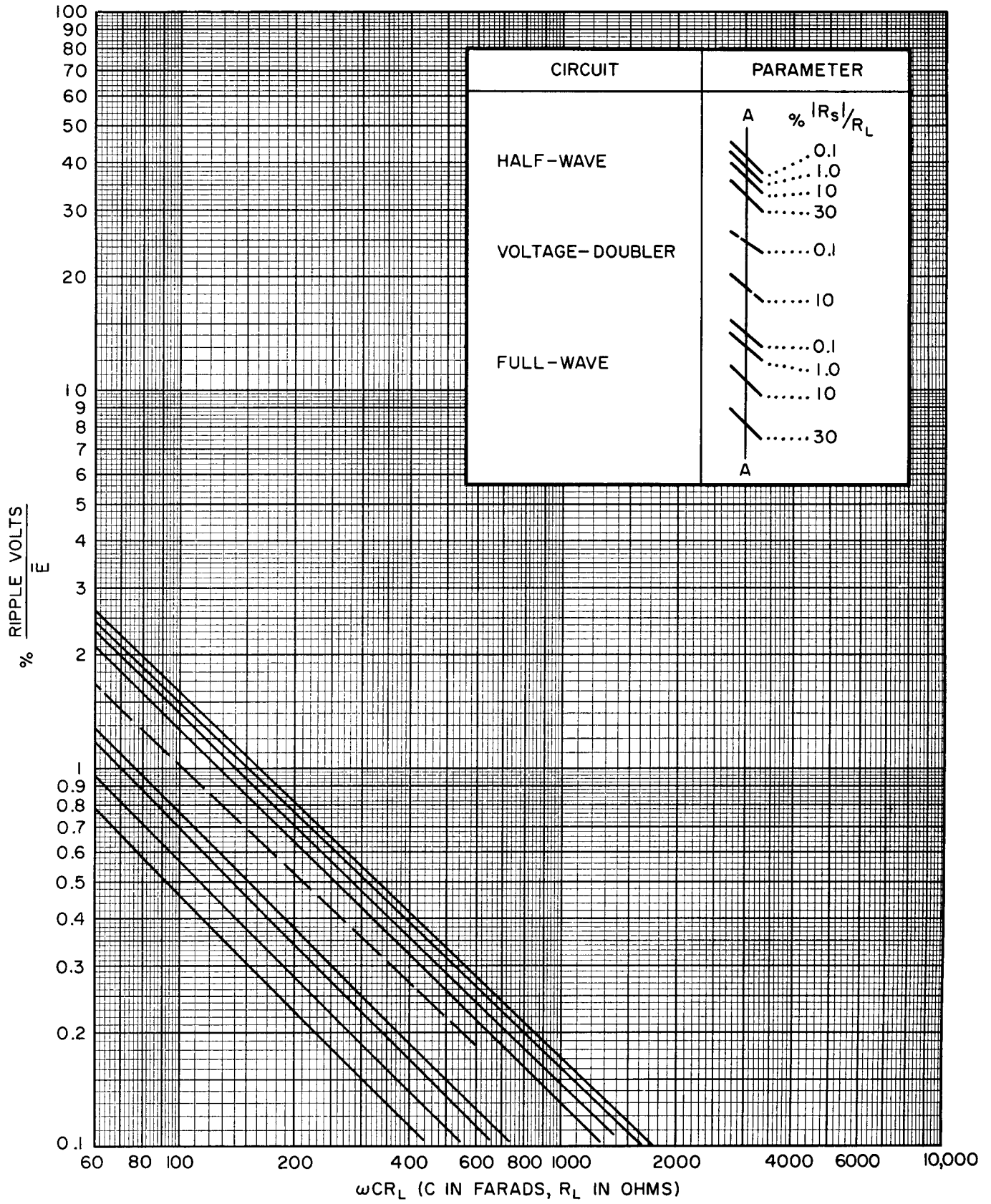


Figure 29. RMS Ripple Voltage for Capacitor-Input-Filtered Load Circuits

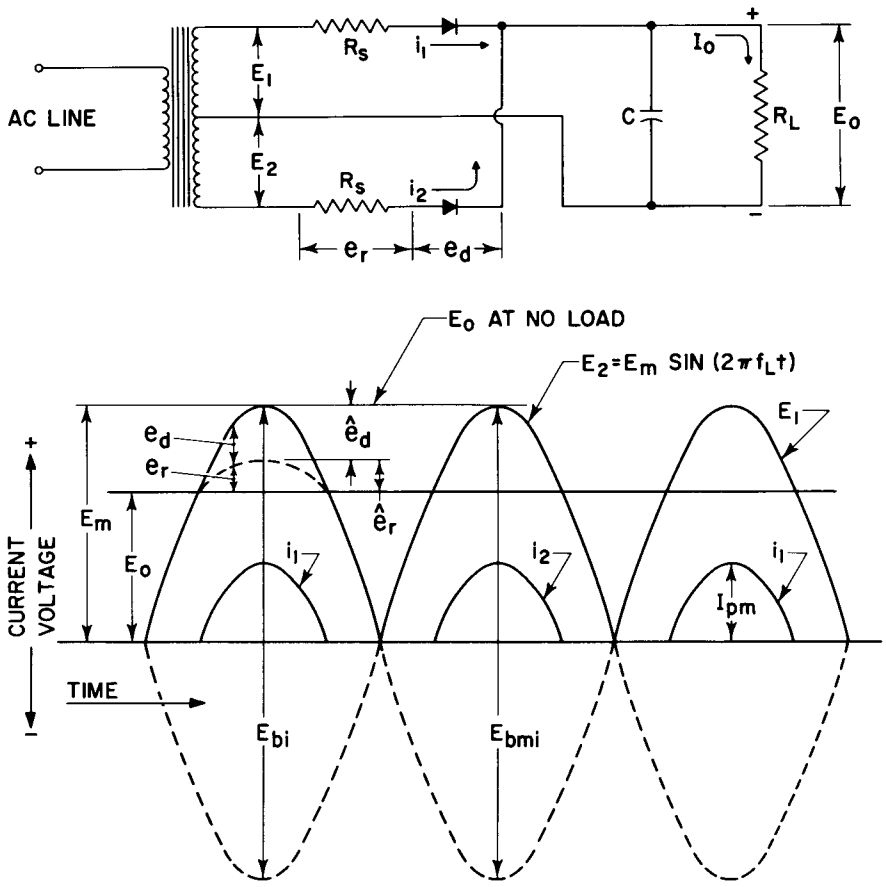


Figure 30. Current and Voltage Waveshapes for Single-Phase, Full-Wave Rectifier with Capacitor-Input Filter of High Capacitance

where E_m = peak anode ac-supply voltage
$\hat{e}_d$ = peak diode-voltage drop for peak current $\hat{I}$ flowing
R_s = resistance in series with the diode

Now, having noted Eq. (25), proceed in the following manner:

(1) Pick an arbitrary value of peak current $\hat{I}$.
(2) From the static diode characteristic, determine the peak diode-voltage drop $\hat{e}_d$ for the current $\hat{I}$ selected.
(3) Compute the dc-output voltage E_o, Eq. (25).
(4) Compute the voltage rectification efficiency η from $\eta = E_o/E_m = E_o/\sqrt{2}E_{rms}$.
(5) From Fig. 32, determine the peak to dc-per-plate ratio and compute the dc current per plate. (Note: $\hat{r}_d = \hat{e}_d/\hat{I}$)
(6) Compute the total dc output current by multiplying the dc-per-plate by the reciprocal of the factor in Table I, row 7 (I_b/I_{av}).
(7) Compute the load resistance, $R_L = E_o/I_o$.
(8) From Fig. 31, determine the minimum filter capacitance. The main ripple frequency f_r is shown in Table I, row 6 (f_r/f).
(9) Obtain the diode rms and maximum surge currents from Figs. 34 and 35; obtain the dissipation of the diode anode and of the added series resistors from Figs. 36 and 37, respectively.

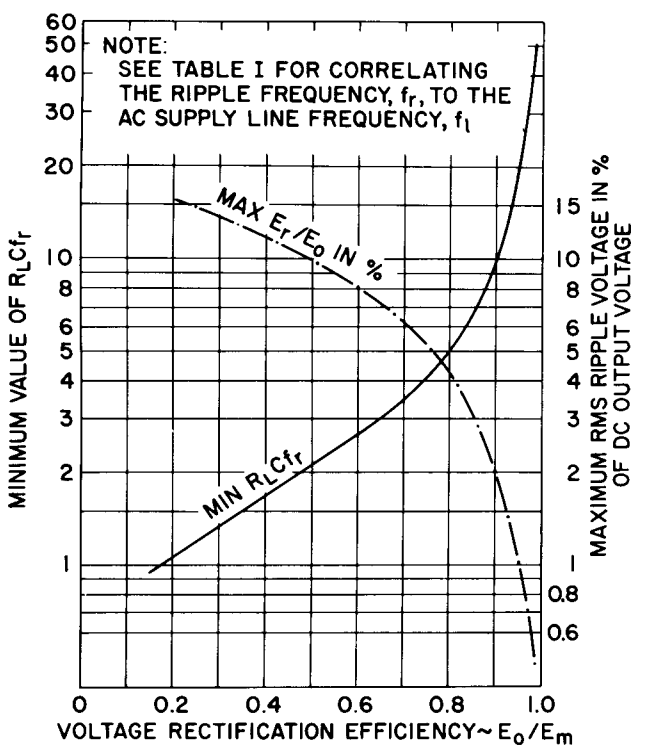


Figure 31. Criteria to Justify the Current Ratios Used in Capacitor-Input Rectifiers of Figs. 32 through 38

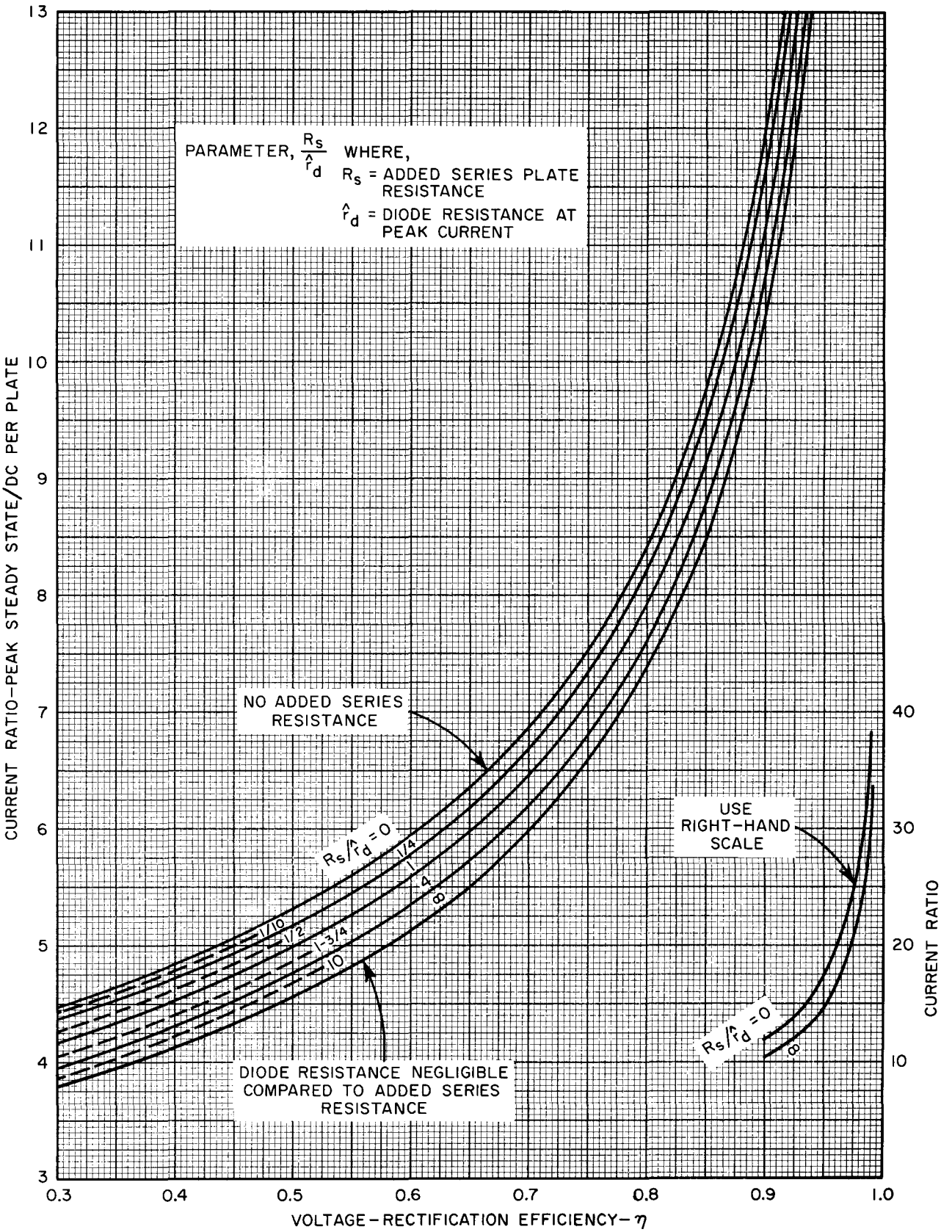


Figure 32. Current Ratio (peak steady-state current to dc-current per plate) vs. Voltage-Rectification Efficiency for Rectifiers with High-Input-Capacitance Filter, Parameter $R_s/\hat{r}_d$

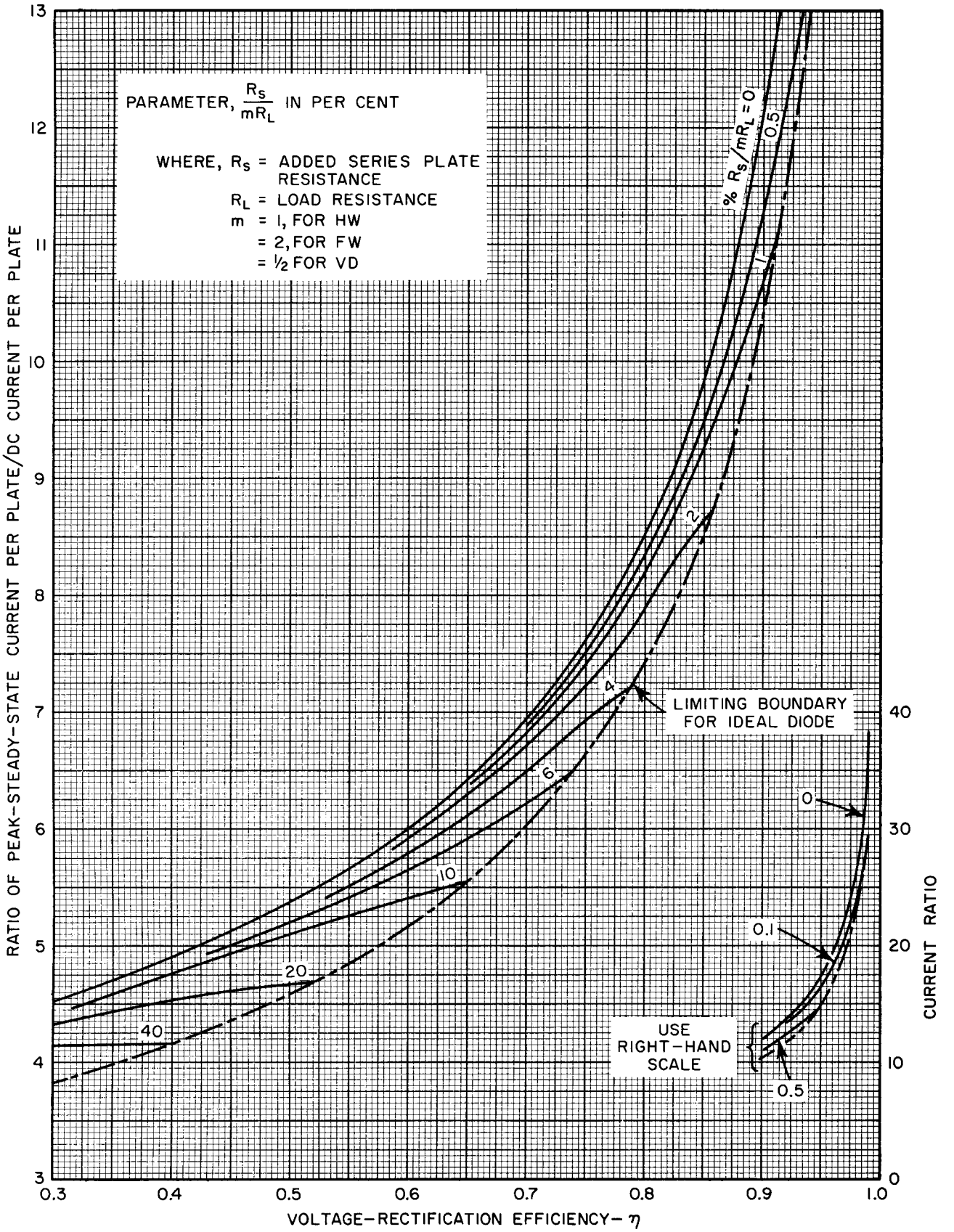


Figure 33. Current Ratio (peak steady-state current to dc-current per plate) vs. Voltage-Rectification Efficiency for Rectifiers with High-Input-Capacitance Filter, Parameter R_S/mR_L

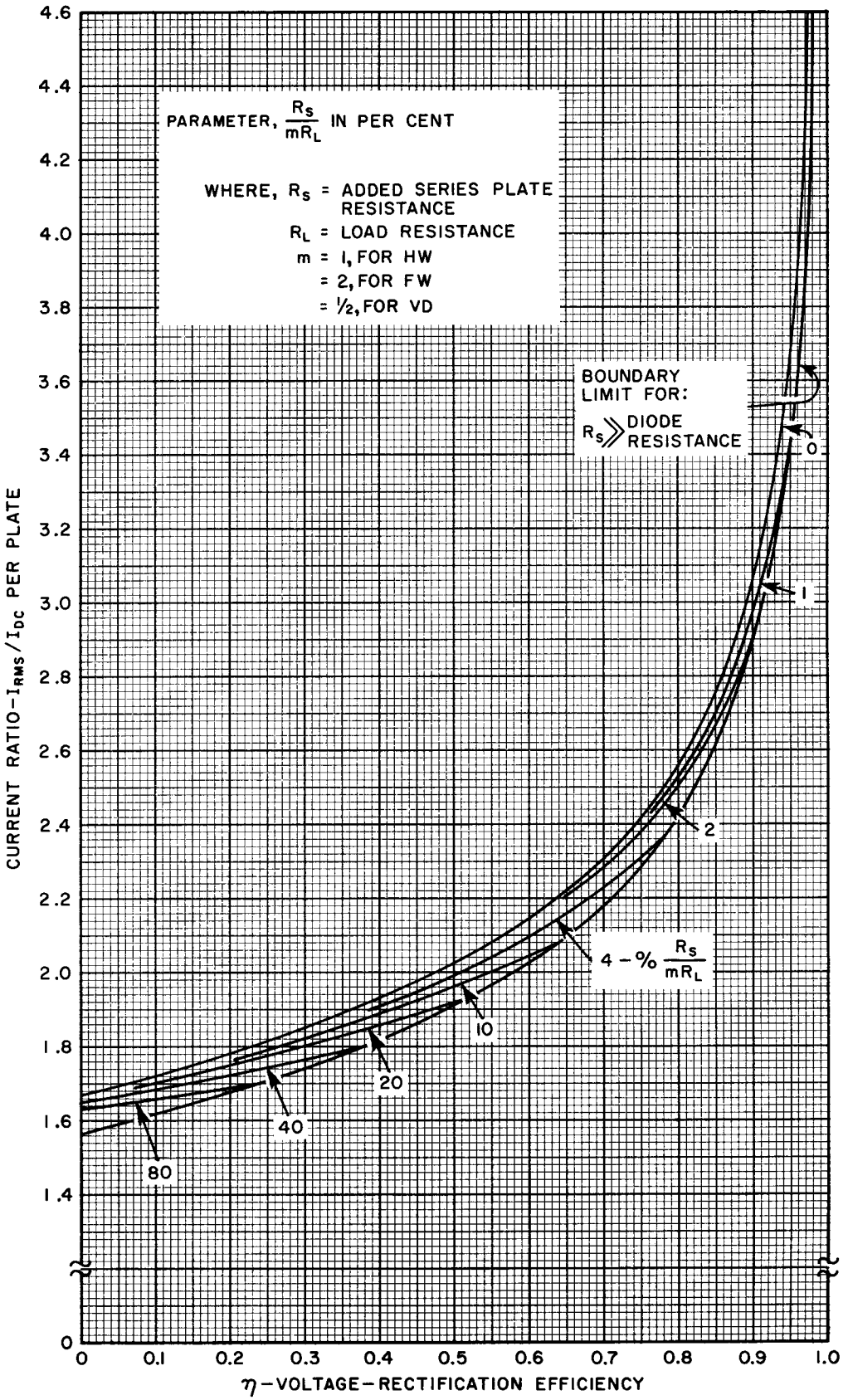


Figure 34. Current Ratio (rms steady-state current to dc-current per plate) vs. Voltage Rectification Efficiency for Rectifiers with High-Input-Capacitance Filter

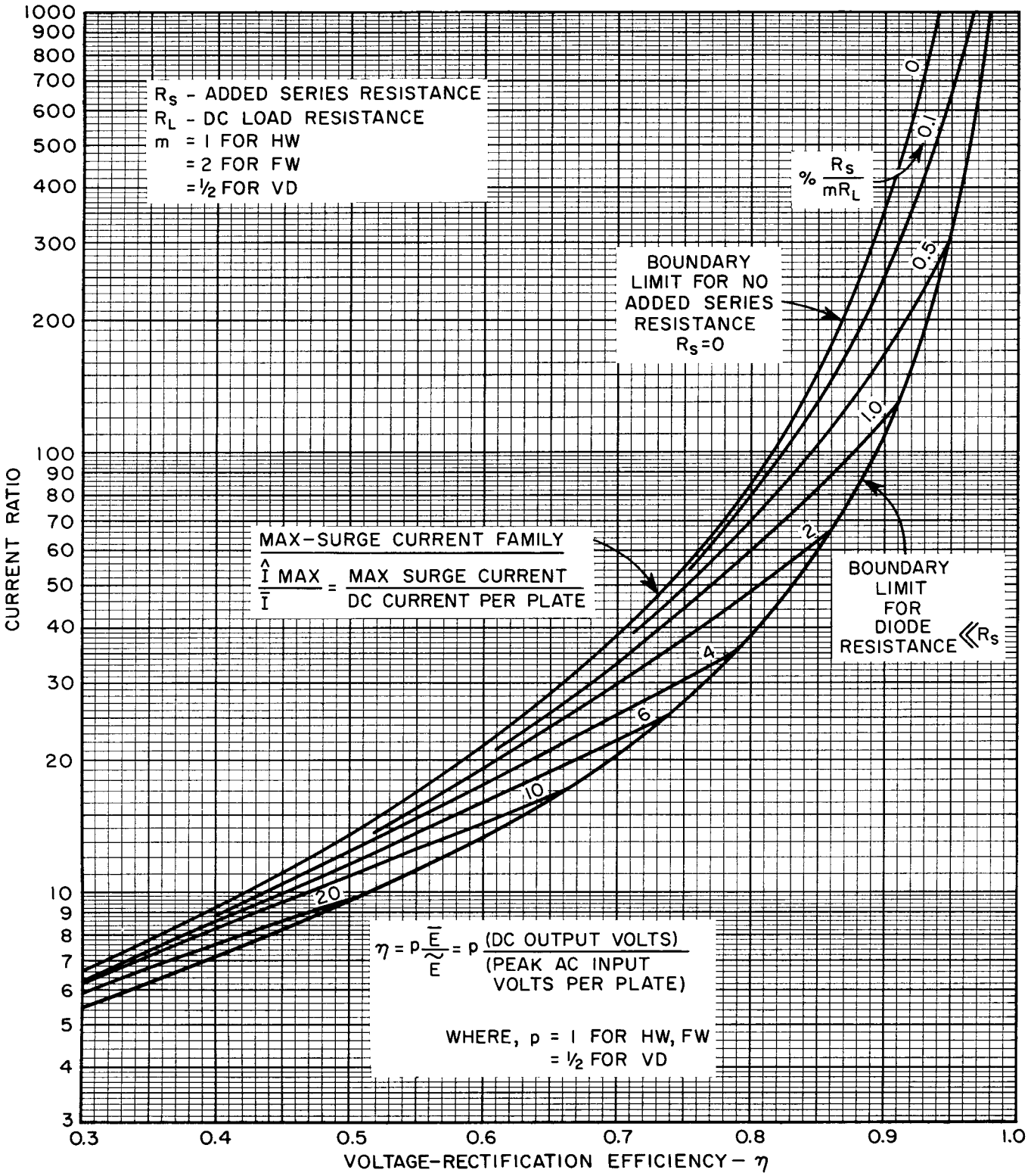


Figure 35. Current Ratio (max. surge current to dc-current per plate) vs. Rectification Efficiency with High-Capacitor-Input Filter

Should the actual capacitance requirement of Step 8 be less than the minimum value, then the assumption that the output voltage is a pure dc value is not justified. A solution can be found for this case from Schade's data, Figs. 25 through 28. However, a series of successive approximations are indicated and, although there are several ways for finding an operation point, the following is a typical procedural example. Only the

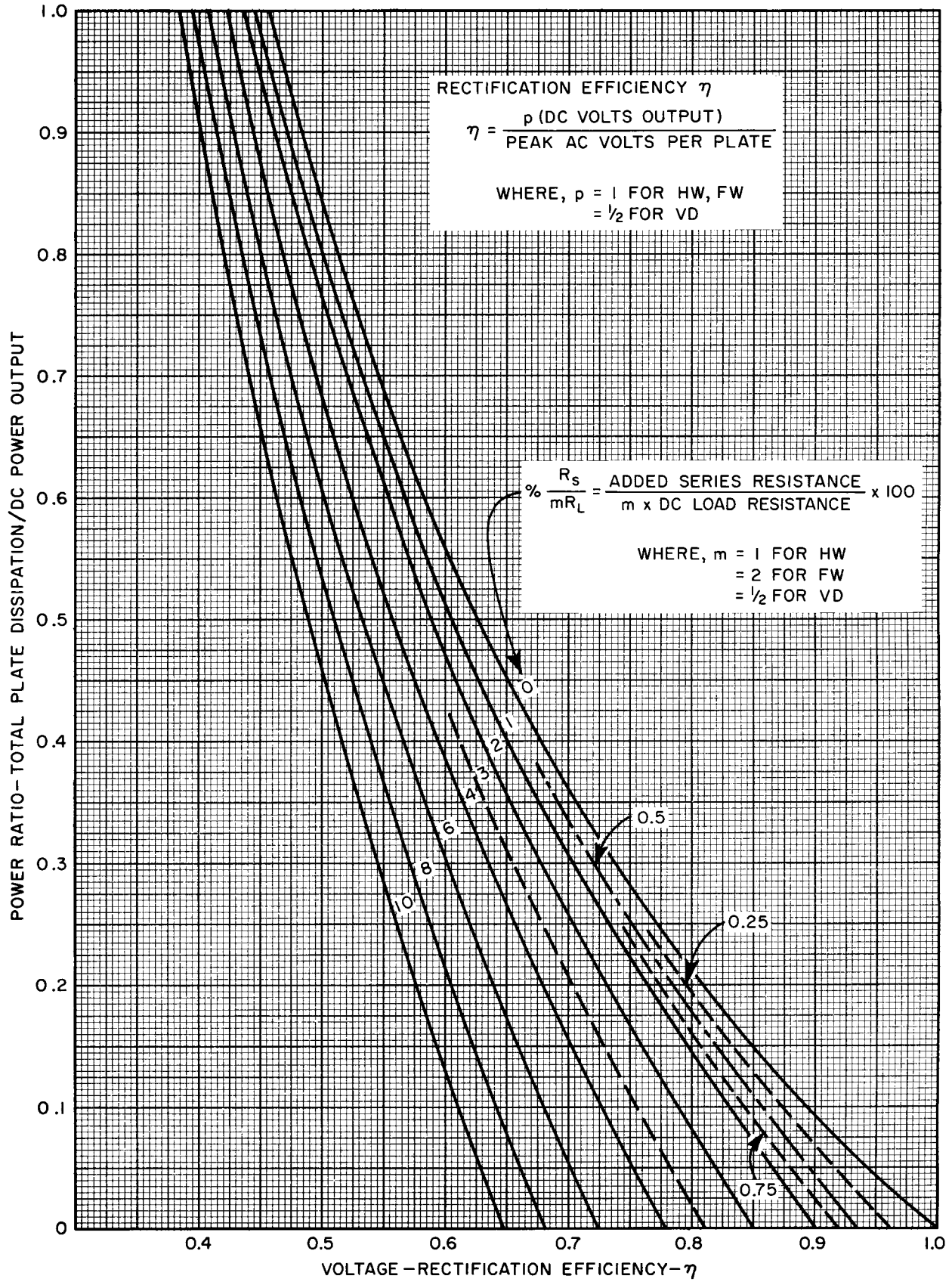


Figure 36. Power Ratio (plate dissipation to dc-power output) vs. Rectification Efficiency for Rectifiers with High-Capacitor-Input Filter; Showing Effect of Adding Series Resistance to Diode

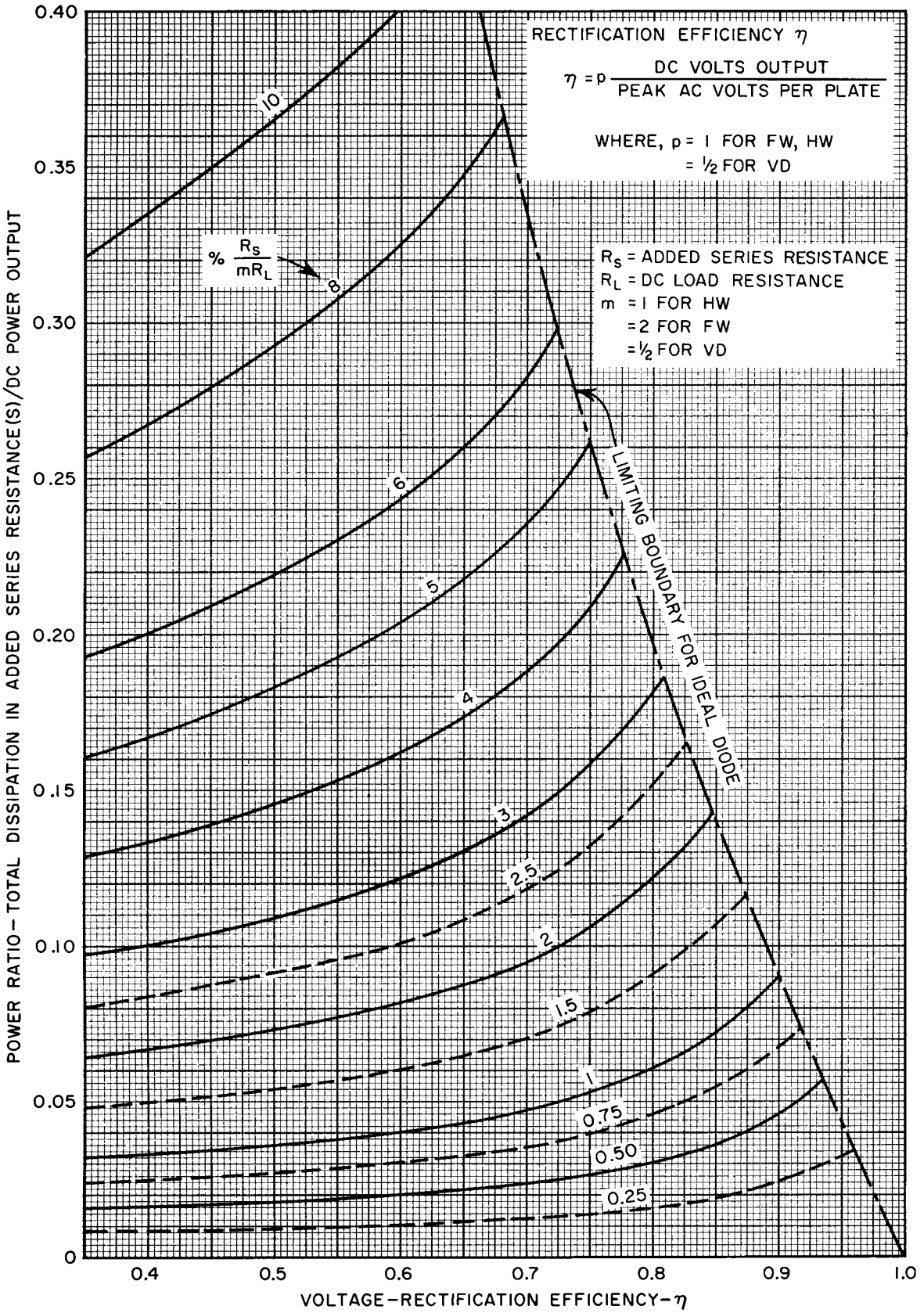


Figure 37. Power Ratio (total dissipation in added series resistance to dc-power output) vs. Rectification Efficiency for Rectifiers with High-Capacitor-Input Filter

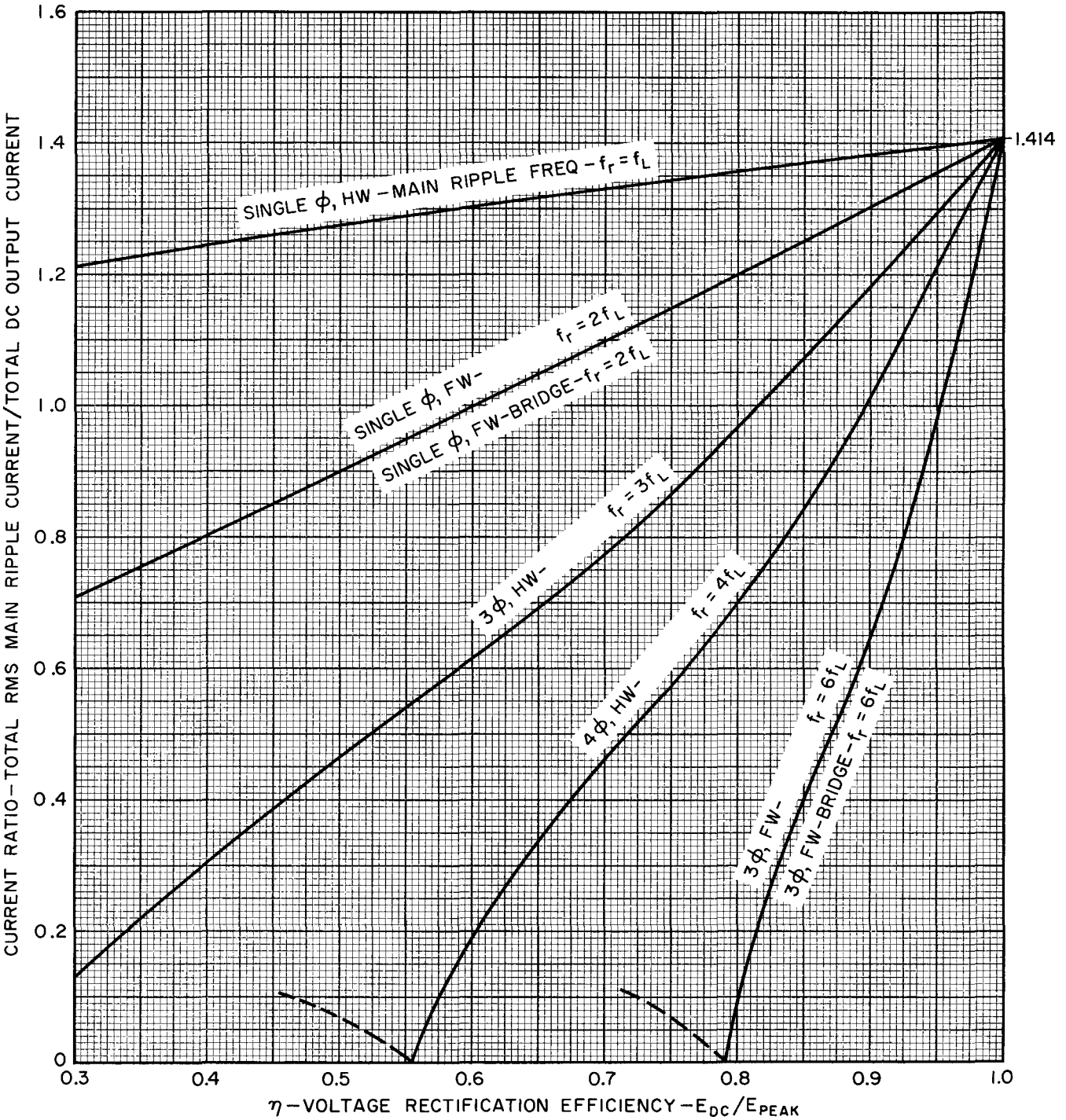


Figure 38. Total RMS Main Ripple Current Characteristics for Rectifiers with High-Capacitor-Input Filter

ac input voltage $\tilde{E}$, the added series resistance R_S, and the static diode characteristic are given. This solution will correlate the peak current, the dc output voltage and current, and the load resistance at one point.

(1) Pick an arbitrary value of peak current $\hat{I}$.
(2) From the static diode characteristic, determine the voltage drop $\hat{e}_d$ for the current $\hat{I}$.
(3) Compute $\hat{r}_d$ from Eq. (24). Determine $\overline{R}_S = R_S + 0.88\hat{r}_d$ (Eq. 24), where R_S is any resistance added in series with the diode.
(4) Assume a load resistance value for R_L.
(5) Compute $\overline{R}_S/R_L$ and $\omega C R_L$ and use these to enter the appropriate Fig. 25, 26, or 27, to get the dc output voltage $\overline{E}$. Note that $\tilde{E}_{max} = \sqrt{2}\,\tilde{E}$.
(6) Compute the dc output current $\overline{I}$ from Steps 5 and 4 ($\overline{I} = \frac{\overline{E}}{R_L}$).
(7) Compute $\hat{R}_S = R_S + \hat{r}_d$ (Eq. 24).
(8) Compute $\hat{R}_S/R_L$ and determine the peak-to-average current-per-plate $\hat{I}/\overline{I}$ from Fig. 28.

(9) Compute the dc output current $\bar{I}$ from Steps 8 and 1.
(10) Compare $\bar{I}$ of Step 6 with $\bar{I}$ of Step 9. If the assumed value of R_L (Step 4) was correct, then both currents should be equal.
(11) If the currents from Steps 6 and 9 are not equal, assume a new value for R_L and repeat Steps 4 through 10 until such an equality is obtained. (Note: If the $\bar{I}$ of Step 9 is larger than the $\bar{I}$ of Step 6, then R_L was assumed too large, and vice versa).

When sufficient observed operating data are available, no assumptions need be made to determine peak or rms diode currents. For example, knowing the dc output voltage $\bar{E}$, the dc output current $\bar{I}$, the added series resistance R_S, and ωCR_L, enter the Schade data for dc output voltage (Fig. 25, 26, or 27) to determine the equivalent dc diode and series resistance $\bar{R}_S$. This $\bar{R}_S$ value is then converted to its corresponding $\hat{R}_S$ and $|R_S|$ values by application of Eq. (24). With these computed equivalent peak and rms resistances, the current ratio curves of Fig. 28 can be used to obtain the peak and rms currents, respectively.

In bridge circuits, each of the charging pulse currents must flow through two diodes connected in series. In Eq. (25), therefore, the peak diode voltage $\hat{e}_d$ should be double that of a single diode. Likewise, in Fig. 32 where the peak diode resistance $\hat{r}_d$ appears in the parameter $R_S/\hat{r}_d$, the value of $\hat{r}_d$ should be double that of a single diode.

Note particularly the three-phase bridge circuit in Fig. 14. In determining the peak current, rms current, and dissipations, use voltage rectification efficiencies only above the value 0.866 (Figs. 31 through 38). Below this value each rectifier passes a current pulse which is no longer a single peak, but becomes double-peaked or M-shaped, and the normalization then does not hold. In addition, the current ratios shown on the ordinates of Figs. 32 through 35 must be halved. This operation is necessary because each diode in this circuit passes two current pulses per cycle of line frequency; the normalizations were based on single pulses.

The diode plate dissipation, exclusive of cathode or filament heater power, in the capacitor-input rectifier circuits may be determined in several ways. A simple relationship, Eq. (26), was derived on the assumption that the diode-current pulse shapes are half-sinusoids and that the instantaneous current through the diode is proportional to its terminal voltage raised to an nth power. The value of n is 1.5 for vacuum thermionic diodes, 2.75 for semiconductors, and zero for gas-filled diodes. By putting the dissipation on a per plate basis, Eq. (26) applies for half-wave, full-wave, single- and multiple-phase, bridge, and voltage-doubler circuits.

$$W_{pp} = K\hat{e}_d I_b \text{ watts per plate} \qquad (26)$$

where $\hat{e}_d$ = peak voltage drop across the diode with peak current flowing
I_b = dc current per plate
K = 0.84, for thermionic vacuum diodes
K = 0.90, for semiconductors
K = 1.00, for gas-filled tubes

In a second method, the diode plate dissipation may be determined from the rms diode current. This current can be evaluated from either Fig. 28 for large filter capacitors, or Fig. 34 for no capacitor restrictions. The dissipation for a single diode will then be

$$W_{pp} = i_{rms}^2 \, |r_d| \text{ watts per plate} \qquad (27)$$

where i_{rms} = diode rms current (See Fig. 28 or 34.)
$|r_d|$ = diode equivalent rms resistance
= $1.07\,\hat{e}_d/\hat{i}_p = 1.07\,\hat{r}_d$ (from Eq. 24)

A third method for determining the diode plate dissipation applies only where a large filter capacitor is used. Here, the dissipation (see Fig. 36) is the total dissipation of all the diodes in the rectifying circuit. This method is particularly convenient when observed data are being taken from an operating circuit.

The resistance dissipation, the dissipation in any resistor R_S in series with each diode in the capacitor-input filter circuits, is given by

$$W_{R_S} = i_{rms}^2 \, R_S \text{ watts} \qquad (28)$$

where i_{rms} is the diode rms current (Fig. 28 or 34).

Instead of adding a series resistor to each diode for protection, it is usual practice, for economy reasons, to add a single resistor of the same magnitude in series with one of the common output load leads. If the single pulses of current through this common output lead have the same amplitude and waveshape as that through a single diode, then the total dissipation in such a resistor will be the value computed from Eq. (28) multiplied by the total number of diodes in the rectifier circuit. This requirement is fulfilled in all the single-phase half-wave, full-wave, and bridge circuits. However, these current pulses will overlap in three-phase circuits operated below certain output voltages. This overlap produces a new waveshape, which the generalized curves do not include. Therefore, for the three-phase half-wave circuit (Fig. 12), this method is not suitable for voltage rectification efficiencies η of 0.500 or less. For the three-phase full-wave circuit (Fig. 13) and for the three-phase bridge circuit (Fig. 14), this method is not suitable below a voltage rectification efficiency η of 0.866.

The total dissipation of a single resistor in series with the common output loading lead, or the sum of the dissipations of separate equal resistors in series with each diode can be obtained from Fig. 37. The total rms current in the common output loading lead is shown in Fig. 38. The same restrictions apply for the three-phase circuits as in the preceding paragraph.

Because typical broadcast or television receiver transformers have equivalent secondary resistances of the order of 100 to 300 ohms, these values cannot be ignored in an accurate analysis of circuit performance. This resistance is part or all of the added series resistance R_S and its magnitude can be determined from the equation

$$R_s = r_{prim} + n^2 r_{sec} \quad (29)$$

where r_{prim} = the dc resistance of the primary winding
r_{sec} = the dc resistance of the secondary winding
n = the primary-secondary step-up ratio, E_{sec}/E_{prim}

Note that in full-wave circuits with a center-tapped transformer secondary, each diode "sees" only half of the total secondary winding and, therefore, in this case, only half of the full secondary resistance and induced voltage apply.

The maximum peak-inverse voltage that can appear across a rectifier with capacitor-input filter occurs at the no-load condition. Using Fig. 30 as a convenient example, all voltages are referred to the center tap of the input transformer (i. e., this tap is also the negative end of the load resistor R_L). The cathodes are both tied to the positive side of the filter capacitor which, at no load, has a maximum positive dc potential equal to the peak of the applied ac voltage $+E_m$. Both anodes are tied to the ac supply terminals and, at the instant when each has peak maximum negative voltage $-E_m$, the maximum peak-inverse voltage E_{bmi} then exists between anode and cathode. This peak-inverse voltage E_{bmi} has a value equal to twice the peak voltage of the ac supply voltage. For bridge rectifier circuits, the ac line connections are "switched" by the rectifier tubes so that the rectifier-tube plates see only positive halves of the ac line voltages. For these circuits, the maximum peak-inverse voltage is equal to the peak of the ac supply voltage. Therefore, for capacitor-input filter circuits at no load, the maximum peak-inverse voltage is

$$E_{bmi} = n\sqrt{2}\ E \text{ volts} \quad (30)$$

where E = the rms ac-supply voltage
n = 2, for half-wave, full-wave, single- and multiple-phase circuits
n = 1, for all bridge circuits

Under loaded conditions for capacitor-input filter circuits, the peak-inverse voltage E_{bi} is

$$E_{bi} = \sqrt{2}\,E + E_o, \text{ for all except bridge circuits} \quad (31a)$$

$$E_{bi} = E_o, \text{ for bridge circuits} \quad (31b)$$

where E_o = the dc-output voltage

A surge or hot-switching transient plate current can flow in the diode of a capacitor-filter rectifier circuit under some operating conditions. A typical but most usual case occurs when a rectifier circuit is switched off after normal operation. After a period of some seconds, if the rectifier circuit is switched on again, a surge current will flow. (Note: Although only the case of one diode is treated, the following discussion applies equally to cases involving more than a single diode.) During the off cycle, the filter capacitor will have had ample time to discharge and, when the circuit is switched on again, this discharged capacitor will momentarily act as a short circuit across the diode. Because the thermionic diode has thermal inertia and does not cool sufficiently, it will supply this short circuit current. The magnitude of this current is many times greater than the steady-state peak current and may cause the diode to sputter or arc. To limit the surge current, it is usual to add a protective resistance in series with the diode. The magnitude of this surge current can be determined from Fig. 35. Fortunately, present-day filter capacitors are no larger than about 100 microfarads, and will charge sufficiently in a few cycles so that the magnitude of the surge current is greatly reduced. The peak surge current can also be computed by assuming a short-circuited load. With all of the peak ac-supply voltage E_m across the diode and a series resistance R_s, then

$$E_m = I_s R_s + \hat{e}_{ds} \text{ volts} \quad (32)$$

and a value of surge current I_s can be determined from the correlated diode voltage drop $\hat{e}_{ds}$ to justify Eq. (32). (Note: Diode data will probably have to be extrapolated because published curves do not usually include currents this large. It is convenient to use log-log graph paper to plot the diode current-voltage characteristic so that these plots will be straight lines.)

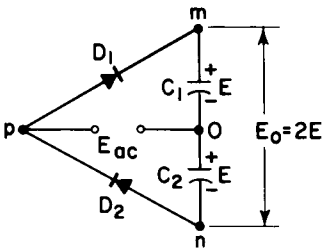
VOLTAGE-MULTIPLYING CIRCUITS[14, 15, 16, 17]

Voltage-multiplying circuits are used widely to obtain a dc high-voltage supply from a lower voltage ac-supply line without the use of a stepup transformer. These circuits are basically a multiplicity of half-wave, capacitor-input-filter circuits. At the various diode terminals, the supply voltage is the ac line voltage with a dc bias voltage superimposed upon it. For loaded outputs in the case of the half-wave and full-wave voltage-doubling circuits, the dc output voltage and diode currents can be determined from Figs. 27, 28, 29, 31, 32, 33, 35, 36, and 37. To obtain the voltage rectification efficiencies η in Figs. 31 to 37, the dc-output voltage should be divided by two to normalize the voltage-doubling effect.

In the full-wave voltage-doubler circuit (Fig. 39,) there are two half-wave rectifiers so connected that their outputs are in series. Capacitor C_1 is charged to the peak ac voltage E when the ac-supply voltage E_{op} is at its negative peak, and the capacitor C_2 is charged to the same peak ac-voltage E when the ac voltage is at its positive peak. The output voltage waveshape is also shown in Fig. 39; this circuit receives its generic name "full-wave" because its output capacitors are recharged twice during a period of line frequency. This action makes the ripple output frequency twice the line frequency, which is an advantage of this circuit over the half-wave doubler. Its disadvantage is that neither of its dc-output load terminals can be "grounded" to the ac line.

In the half-wave voltage-doubler circuit (Fig. 40), the voltage-doubling process can be more readily seen if we replace the capacitor C_1 by a battery whose voltage is equal to the peak ac-line voltage E. The voltage E_{pm}, from this battery in series with the ac-line voltage, is a displaced sine-wave which has a positive peak value of 2E with respect to the point p. This voltage

source charges the capacitor C_2 through the diode D_2 to plus 2E volts. With C_1 replaced into the circuit, it will charge through the diode D_1 to the voltage E when the ac-line voltage is most negative with respect to the point p. The output capacitor C_2 is charged once during a cycle of line frequency, hence its generic name "half-wave." In contrast to the full-wave voltage-doubler circuit, one side of the dc output is common to one side of the ac line, an advantage that is offset by a lower ripple frequency in its output.



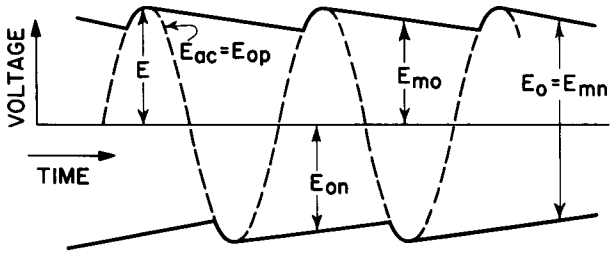
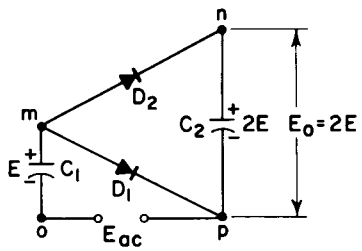


Figure 39. Full-Wave Voltage-Doubler Rectifier and Voltage Waveshapes

The half-wave voltage-tripler circuit (Fig. 41) is derived from the half-wave doubler circuit (Fig. 40) by adding the branch consisting of the capacitor C_3 and the diode D_3. The capacitor C_3 will charge to the voltage 2E when the ac-line voltage is at its negative peak with point p as a reference. The points m and p will then be at zero potential and capacitor C_2 will then charge capacitor C_3 to the voltage 2E. The charging current must flow through diode D_3, the capacitor C_1, the ac-supply line source impedance, and C_2.

Another voltage tripler circuit is found in the generalized voltage multiplying circuit of Fig. 43 in that part of the circuit included within the double lines. It consists of a half-wave voltage-doubler circuit (as in Fig. 40) with a single half-wave rectifier added. Because the capacitor C_1 is charged on the opposite half-cycle that C_2 and C_3 are charged, this circuit might be called a "full-wave" circuit. However, there will still be some ripple output at line frequency because the circuit is not balanced. Two-thirds of the voltage is maintained during the first half-cycle of line frequency and only one-third on the other half-cycle.

The half-wave voltage quadrupler circuit shown in Fig. 41 is derived by adding another step to the half-wave voltage-tripler circuit and demonstrates that any desired voltage multiplication can be achieved by similar additions.



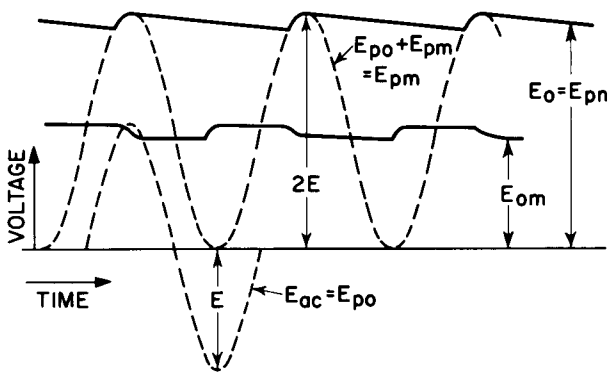


Figure 40. Half-Wave Voltage-Doubler Rectifier and Voltage Waveshapes

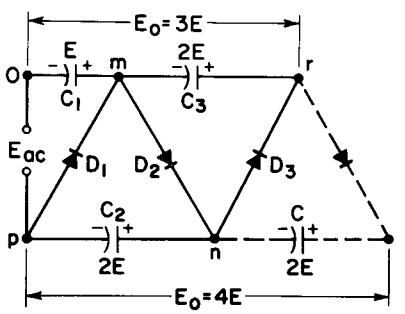


Figure 41. Half-Wave Voltage-Tripler (and half-wave voltage-quadrupler, by addition of dotted section) Rectifier

The full-wave voltage-quadrupler circuit of Fig. 42 is derived from two half-wave voltage doubler circuits back-to-back. (Compare with Fig. 40.)

The voltage quadrupler and sextupler of Fig. 42 and the generalized voltage multiplying circuit of Fig. 43 contain all possible voltage multiplying circuits. The latter is asymmetrical about the ac line terminals while the former is symmetrical. Note the difference in the diode connections in these two circuits; in both diagrams they are connected in series except for the single break in the asymmetrical case.

The maximum peak inverse voltage across any diode

in all of these multiplier circuits is twice the peak of the ac-line voltage 2E.

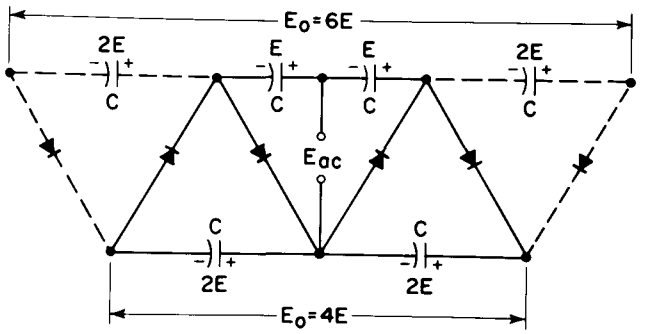


Figure 42. Full-Wave Voltage-Quadrupler (and full-wave voltage-sextupler, by addition of two dotted sections) Rectifier

There is no advantage in using capacitors of differing capacitance values in these circuits, but they should have as high a value as possible to reduce the ripple output and to preserve the output voltage regulation.

HIGH-VOLTAGE RECTIFIER TUBE DESIGN

The high-voltage, low-current, dc-supply needed for cathode-ray tubes is usually obtained from either a radio-frequency source[18, 19] or, as in most TV receivers, from the horizontal output transformer's stepped-up high-voltage pulse during the retrace or kick-back period.[20, 23, 24] Special design considerations must be given the associated rectifier tubes mainly because they are subjected to peak-inverse voltages as high as 20 to 40 kilovolts.

Potential Gradients and Field Emission

It is well known that the forward voltage drop of television high-voltage rectifiers is quite small (in the order of hundreds of volts) during the duty cycle of the rectifier. However, the inverse voltage may be as high as 40,000 volts and more when the anode voltage is negative with respect to the cathode, because the value of the inverse voltage is substantially two times the rectified dc output voltage for a sinusoidal voltage source and approximately 1.2 times the rectified dc output voltage in unidirectional pulse supplies. The electrode structure is subject to very little electrostatic stress during the duty cycle (anode positive) while high electrostatic potentials exist between anode and cathode during the inverse voltage cycle (anode negative). It is known that, with a sufficiently high positive potential gradient at a cold metal surface, electrons will be emitted. This emission, termed "field emission," occurs at smooth negative surfaces, when the potential field gradient is in the order of 100 kilovolts/cm (250 kilovolts/inch, 250 volts/mil). Because the field gradient is increased at a point projecting from a surface or at a sharp edge, field emission may occur when the normal surface gradient is 50 kilovolts/cm or less, depending on the edge sharpness or the length of points due to surface irregularities. High-potential fields occur in the rectifier tube only during the inverse voltage cycle, and only when the anode is negative. The occurrence of field emission depends, therefore, on the then positive potential gradients at the anode surface.

Field emission is most undesirable because electrons leaving the anode will bombard the cathode at tens of kilovolts equivalent electron velocity, destroying its emitting surface and producing soft X-rays. Those electrons which miss the cathode and can eventually strike the glass bulb can cause electrolysis and glass disintegration (see the following section on Secondary Emission).

Precautions to prevent cold field emission consist of removing all small surface irregularities by electropolishing. Sharp edges and corners cannot be tolerated, and should be rolled to give a relatively large radius of curvature. Small points such as those caused by oxide particles evaporated on the anode during cathode processing can be burnt off by spot knocking, i. e., by temporarily applying approximately a twice-normal inverse potential to the rectifier tube so that the extremely high potential gradient causes a high field emission current density from the particles and causes them to evaporate.

Some idea of the gradients inside a diode can be had if it is assumed that the structures are concentric cylinders. The anode structure is such a cylinder but, as shown in Fig. 44, the cathode structures only roughly approximate cylinders. The voltage V between two concentric cylinders at any radius r is

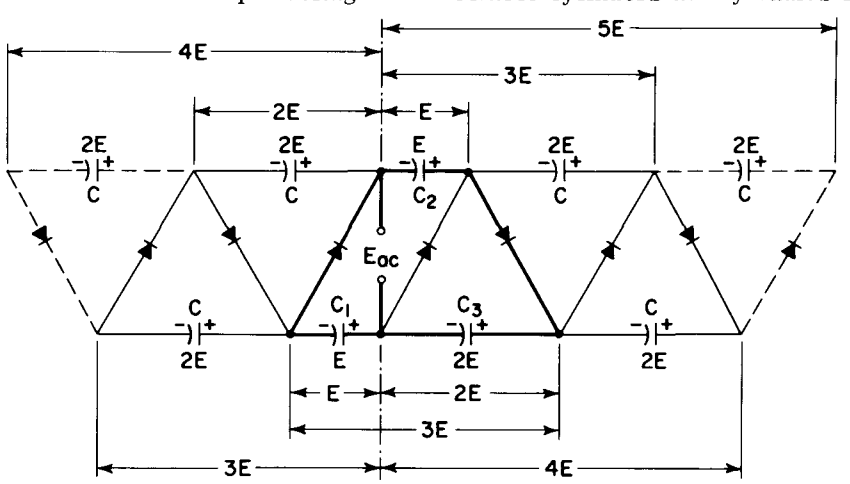


Figure 43. Half-Wave Voltage-Multiplying Rectifier

$$V = V_{max} \frac{\ln r / r_1}{\ln r_2/r_1} \text{ kilovolts} \qquad (33)$$

and the potential gradient $E = dV/dr$ at any radius r is

$$E = \frac{V_{max}}{r} \frac{1}{\ln r_2/r_1} \text{ kilovolts/cm} \qquad (34)$$

where V_{max} is the inverse voltage in kilovolts between the outer cylinder (anode) with radius r_2 and the inner cylinder (cathode) with radius r_1, and r is any radius between the limits r_1 and r_2 measured in centimeters. The logarithm ln is the natural log of base ϵ.

From Eq. (34) it is apparent for r equals r_2 that the anode radius should be made as large as possible to minimize the anode field gradient. Although minimum field gradient will occur at the cathode when $\ln r_2/r_1$ equals unity (i.e., the ratio $r_2/r_1 = 1/\epsilon = 0.368$), a preferred choice of cathode radius r_1 is probably nearer the value $r_1 = 0.1\ r_2$. Under the latter condition the cathode gradient is doubled, but this undesired result is more than offset by a reduction of the anode gradient to half that for conditions of minimum cathode field gradient, (see subsequent discussion of Potential Gradient at Cathode).

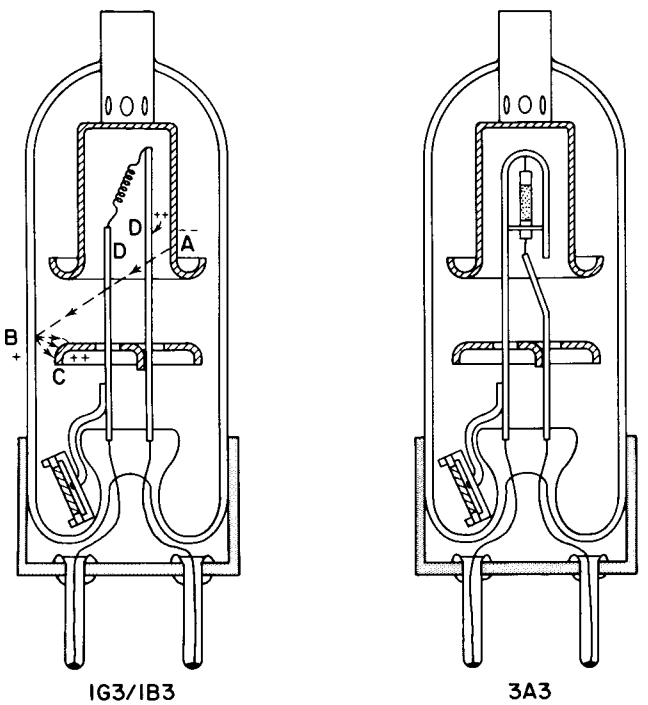


Figure 44. Cross-Sectional Views of 1B3 and 3A3 High-Voltage Rectifier Tubes

Secondary Emission

The presence of secondary emission in the high-voltage rectifier tubes is perhaps even more intolerable than the cold field emission. It is known that many insulators, especially the metallic oxides, are prolific secondary emitters. It is therefore not too surprising to find that many troubles can be tracked down to secondary-emission effects on the inside glass walls of the containing envelope. For example, look at the 1B3 structure in Fig. 44. Conditions are assumed at the time when the anode has its peak negative voltage with respect to the cathode. Electrons will leave A on the anode due to the high positive field gradient. These electrons will accelerate towards the cathode supports D, D, and the shield F. Those electrons not collected will continue towards the glass tube wall and, when they strike, either of two things will happen. If the ratio of secondary electrons released to primary electrons striking is greater than unity, then point B will become charged more and more positive until the leaving secondary electrons can just be collected by the slightly more positive shield F. Observe that there is a constant current flowing (for the momentary conditions assumed) and the primary current is dissipating considerable heat by striking the glass at high voltage. Eventually, the glass will show signs of electrolysis in such a region; the tube may become gassy or the glass wall may crack due to heat strains. On the other hand, if the ratio of secondary electrons released to primary electrons striking the glass wall is less than unity, then point B will collect electron negative charges until it has become so negative that it will turn almost all the primaries back and no glass disintegration will occur.

It is very important that the cathode be placed deep enough inside the anode cup so that, during cathode breakdown, materials cannot spatter on the glass walls. The oxides of barium, strontium, and calcium are all high secondary-ratio emitters. Furthermore, Deckert[21] and Gallup[22] have reported that, by using lead bulbs fired in hydrogen for 10 minutes at 300 C, a layer of lead hydride is produced which apparently gives uniform secondary emission ratios less than unity because the "charging up" and the formation of "hot spots" on the glass bulb walls have been eliminated.

Potential Gradient at the Cathode

Although the maximum potential gradient at the cathode surface is greater than that at the anode surface, see Eq. (34), it is of the opposite sign and, therefore, cannot pull electrons from the cathode when the peak-inverse voltage is applied. However, some restraint should be put on the magnitude of this maximum gradient at the cathode. This gradient is responsible for a mechanical force on the cathode coating which tends to pull the oxide coating from its base material. This force is proportional to the square of the field gradient and is

$$F = 4.5 \times 10^{-4}\ E^2 \text{ grams/cm}^2 \text{ area} \qquad (35)$$

where E is the field gradient in kilovolts per centimeter

Where construction is not symmetrical (such as in the 1B3, Fig. 44), a differential force results which may move the whole cathode, as was reported by Deckert,[21] whose motion pictures showed the mechanism for producing "melted-lead" filament failures. First, the filament would vibrate and loosen particles of its oxide coating; these particles, in turn, were pulled

to the plate by electrostatic force. Next, these particles would become white hot due to forward electron bombardment. On the reverse voltage cycle, these white-hot particles would emit electrons which were collected at the filament support leads at many kilovolts. The support rods heated by this reverse emission would get red hot and eventually melt. The solution in this particular case was to change to a stiffer filament and use a cathode coating, such as carbonyl, which adhered better to the base material. A symmetrical structure, as perfectly centered as possible (see the 3A3 structure, Fig. 44) should prevent the cathode vibration. From this standpoint, the new nuvistor with concentric cylinder construction is ideal.

Electrolysis

Glass electrolysis is due to the ionic conduction which is always present to some degree in glass with large potential gradients. It is caused by ionic currents which result when the negative sodium ions and positive oxygen ions move in opposite directions in a given field gradient. The magnitude of this ionic current is temperature sensitive and increases rapidly with the glass temperature; it is proportional to the voltage between leads, is greatest at the moment of peak inverse voltage, and is dependent on the glass chemical composition. Electrolysis causes tube failure when the chemically active sodium ions collecting at the anode lead (because it becomes most negative at the time of maximum peak inverse voltage) reacts to destroy the glass-metal bond. In lead glasses, there is a discoloration in the glass in the region where the sodium ion has given up its charge, and the resulting active sodium atom changes the lead oxide to metallic lead. The oxygen ion collects at the positive or cathode lead and will eventually break the bond between the stem lead and the glass, although there usually is no discoloration visible from this charge collection. It is interesting to note[21] that the silver bulb etching ink on the 1B3 was changed to a non-conducting ink when heavier electrolysis was always observed on the side of the bulb having the reduced electrical resistance path due to the conducting ink. Lead-glass bulbs, replacing the older lime-glass bulbs, are greatly superior for reducing electrolysis.

X-Rays

Whenever a beam of electrons strikes any target with velocities equivalent to thousands of volts, the presence of X-rays should be investigated. The cold emission currents, arising during the 40-kilovolt spot knocking of the 1B3, produced sufficient X-ray radiation to require tube modifications. The stem shields and getter assembly had sharp points minimized, but the biggest improvement in reducing X-ray radiation was obtained when lead glass bulbs were introduced to reduce electrolysis and secondary emission.

Diode Currents

The relatively high frequency used in rf supplies and the high repetition rates of television pulsed supplies permit the use of simple filter circuits. Mautner and Schade[19] show that, if the discharge-time constant of the output filter is long compared to a period of rf or to the time between pulses ($\omega CR \geq 500$), then the circuit operation is comparatively simple. The output voltage can be considered constant, and will have the value KE_m (where K is the voltage rectification efficiency) for a given load. This relation is shown graphically in Fig. 45.

The diode conduction angle can be derived from

$$\psi^o = 2 \cos^{-1} K \tag{36}$$

Figs. 45A and 45B show the relation between conduction time and diode peak $\hat{I}$ and average $\bar{I}$ currents. If the peak current is assumed to be constant during the conduction period (shown as dotted), then the equivalent average dc output current $\bar{I}$ is determined by ratio of the total time for a repeating cycle to the conduction time. If the assumed rectangular conduction current is multiplied by the factor $\pi/2$, it will give the peak current for a sinusoidal diode current, a value which is a sufficiently close approximation for all practical purposes. Hence, the ratio of peak to average dc output current, for rf supplies, is

$$\frac{\hat{I}}{\bar{I}} = \frac{\pi}{2} \frac{360^o}{\psi^o} = \frac{\pi}{2} \frac{90^o}{\cos^{-1} K} \tag{37}$$

and, for pulsed voltage supplies, is

$$\frac{\hat{I}}{\bar{I}} = \frac{\pi}{2} \frac{180^o T_H}{\psi^o T_R} = \pi \frac{45^o}{\cos^{-1} K} \frac{T_H}{T_R} \tag{38}$$

where ψ is the conduction angle, K is the voltage rectification efficiency E_{dc}/E_m, and the ratio T_H/T_R is the ratio of the total time of one cycle to the time duration of the voltage surge pulse.

When no load current is drawn, the available dc output voltage can have a maximum value equal to the peak of the rf or pulse voltage E_m. However, when load current is drawn, the resulting drop in the source impedance and/or series resistance including the rectifier tube drop will give output voltage below this maximum.

The relationship between the voltage rectification efficiency K and the ratio of the dc load resistance R_L to the series source impedance Z_s, for rf supplies, is

$$\frac{R_L}{Z_s} = \left[\frac{1}{\frac{1}{K} - 1}\right] \cdot \frac{2 \times 360^o}{\psi^o} = \left[\frac{1}{\frac{1}{K} - 1}\right] \cdot \frac{360^o}{\cos^{-1} K} \tag{39}$$

and, for pulsed voltage supplies, is

$$\frac{R_L}{Z_3} = \left[\frac{1}{\frac{1}{K} - 1}\right] \cdot \frac{2 \times 180^o}{\psi^o} \cdot \frac{T_H}{T_R} = \left[\frac{1}{\frac{1}{K} - 1}\right] \cdot \frac{180^o}{\cos^{-1} K} \cdot \frac{T_H}{T_R} \tag{40}$$

where T_H/T_R has the same significance as explained in Eq. (38), and Z_s is the surge impedance $\sqrt{L/C}$ of the source or, if sufficiently large, the resistance R_s in series with and including the diode equivalent resistance.

In the Murakami reference,[24] practical horizontal output transformers for black-and-white television receivers have a retrace time T_R of approximately 9 microseconds; and for color television receivers have a retrace time T_R of approximately 13 microseconds. In the United States, present standards require the total period T_H between horizontal synchronizing pulses to be 63.5 microseconds. With the above information and as assumed voltage-rectification efficiency of K = 0.90, $\psi^o/2 = \cos^{-1} K = 25.9^o$, values for Table II were prepared.

Table II

Typical Current Ratios for
Kinescope High-Voltage Rectifiers

Ratio	Pulsed Supply for		RF Supply
	Black-and-White Receiver	Color	
$\hat{I}/\bar{I}$	38.5	26.6	11.05
R_L/Z_s	441.	305.	125.

DAMPER-TUBE RECTIFIER DESIGN

The operation of the damper tube in typical television circuits has been thoroughly analyzed by Schade[20] and Murakami.[24] The plate dissipation of the damper tube (and also the horizontal-deflection output tube) can be determined from oscilloscope current and voltage readings as has been demonstrated by Angel.[25]

In the design of damper diode tubes, such as the 6DE4 and the 6AU4-GTA, consideration must be given to the insulation requirements of the 5-kilovolt maximum peak inverse-voltage rating. Not only does this voltage appear between anode and cathode but, because the heater may be run at ground potential, the 5 kilovolts will also appear between heater and cathode. Added to the high voltage problems, are the sputtering at cathode and anode surfaces and the plate dissipations and temperatures which stem from the high ratings for peak current and average current, 1.3 amperes and 200 milliamperes, respectively. In the following list of design considerations, the first four items represent recent improvements which have greatly improved the quality and life of damper tubes:

1. Additional heater-cathode insulation has been added by surrounding the regular folded heater with a tungsten spiral which has low turns per inch and is heavily coated with alumina spray. (See Fig. 46B.) This helix is prevented from slipping out of position by staking one end of it to the bottom mica.

2. Use of an improved cathode-emission coating has resulted in a great improvement in quality and life-test performance. The new coating consists of 16 per cent of nickel powder added to a regular carbonate coating; the resulting mixture is sprayed to an apparent density of 1.3 milligrams per cubic centimeter. This coating does not peel off the cathode and does not arc and sputter as readily as the conventional coating material; it is known as "carbonyl nickel" spray coating. With this coating, direct current densities of 100 milliamperes per square centimeter have been obtained with good life.

3. The anode construction has several salient improvements. The seven-mil thick material is a "sandwich" of shiny nickel on the inner side, facing the cathode. The outside of the sandwich is an iron-aluminum alloy which is oxidized during exhaust to give a black satin surface which is a good heat radiator. The older carbonized plate materials were always a source of sputter due to the presence of loose carbon particles. The anode structure has four louvered port holes (see Fig. 46A) which keep gas pressures low during breakdown and during any arcing while tubes are being processed. Keeping the pressure low at all times during the tube manufacture is a decisive factor in producing good, stable emitting surfaces.

4. The top and bottom mica insulators have wide 50-mil slots, semicircular in shape and centered about each of the four anode supports (see Fig. 46A). The distance from the mica to the plate barrel was spaced to give a gradient of 65 volts per mil between the mica and anode barrel. This result was accomplished by undercutting the plate structure as shown in Fig. 46A. These precautions remove excessive gradients from the mica and tend to remove undesirable creeping discharges along the mica surfaces. Any mica surface leakage developed during processing is removed by application of a voltage of 10 to 12 kilovolts from a spot knocker.

5. Anode-to-cathode spacing is approximately 30 mils; this spacing was chosen to give peak inverse voltage gradients of 165 volts per mil. Because cold emission discharge starts at about 250 volts per mil, a design maximum of 150 to 180 volts per mil should never be exceeded. Cold discharge from the top and bottom of the sharp edges of the anode barrel is prevented by making these parts with turned edges that give "rounded corners" to the field in this region. It is good design to make the radius of such a corner equal to the spacing between anode and cathode.

6. If there is need to increase the ratings of these damper tubes so that more heat dissipation would be required, several further improvements are possible. The use of a plate material having substantial copper core and the use of internal anode leads of nickel-clad copper (Kool grid) would permit the removal of considerably more heat than the present 6 to 6-1/2 watts. In addition, the getter, now flashed so that one side of the anode cannot "see out" of its opposing glass wall, could be removed to the dome of a longer glass envelope. This would improve the heat radiation efficiency but would require some provision for shielding the top mica from getter-flash leakage.

7. The present tubes are limited to their present maximum peak inverse voltage of 5000 volts by the fact that they are single ended; the stem structures will not take much more. A higher voltage require-

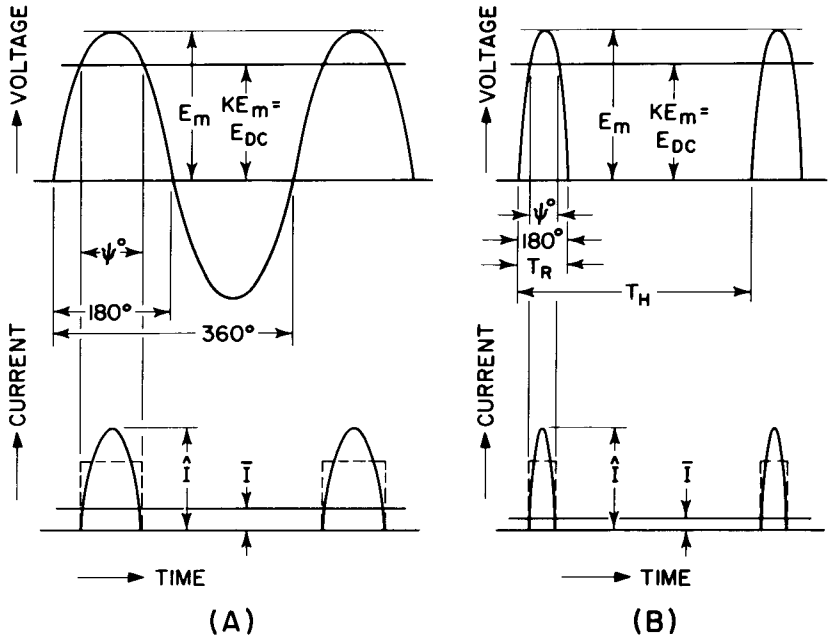


Figure 45. Diode Conduction: (A) RF Supply, (B) Pulsed Supply

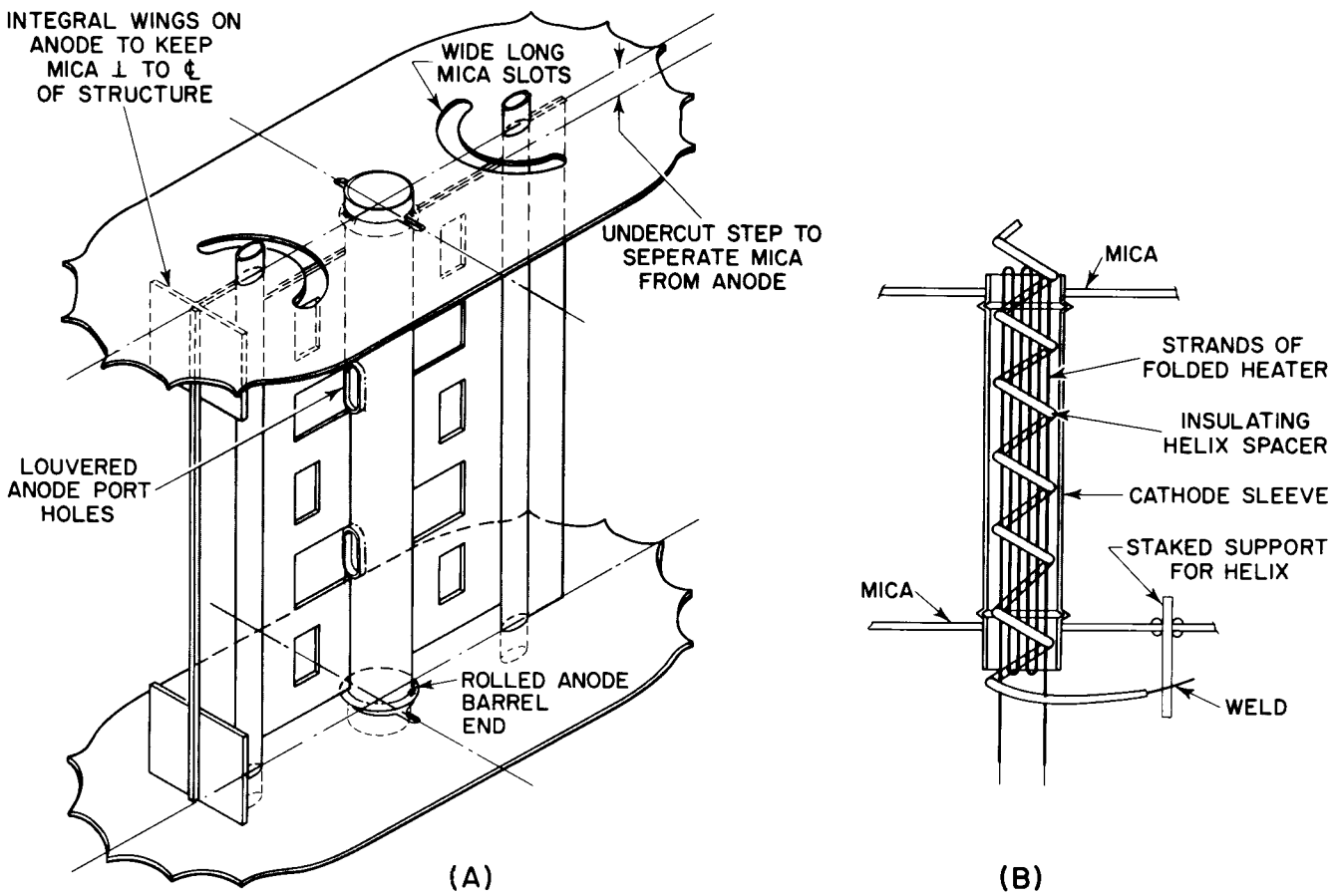


Figure 46. Constructional Details of Damper Diode (6DE4 and 6AU4-GTA): (A) plate and mica assembly, (B) cutaway cathode showing position of helix-insulating spacer and heater

ment would probably be best handled by taking the anode lead out through a top cap.

REFERENCES

1. Langmuir, I., "Electrical Discharges in Gases, Part II, Fundamental Phenomena in Electrical Discharges," Rev. of Mod. Phys., Vol. 3, No. 2, pp. 247-248, April 1931
2. Kusunose, Y., "Calculation of Characteristics and Design of Triodes," Proc. IRE, Vol. 17, No. 10, pp. 1706-1749, Oct. 1929
3. Prince, D. C., and F. B. Vodges, Mercury-Arc Rectifiers and Their Circuits, McGraw-Hill, New York, 1927
4. Armstrong, R.W.,"Polyphase Rectification, Special Connections," Proc. IRE, Vol. 19, p. 78, 1931
5. Schade, O.H., "Analysis of Rectifier Operation," Proc. IRE, Vol. 31, pp. 341-361, 1943; "Electron Tubes, Vol. II," RCA Review, Radio Corporation of America, Princeton, N.J., 1949
6. Heymann, F. G., "Rectifier Filter Characteristics," Wireless Eng. (Brit.), Vol. 32, No. 6, pp. 147-154, July 1955
7. Barbrow, L.E., and J. F. Meyer, "Characteristics of Tungsten Lamps," Bureau of Standards Jo. of Research, (U.S. Dept. of Commerce), No. 9, pp. 724-729, July-Dec. 1932
8. Langmuir, I., and H.A. Jones, "The Characteristics of Tungsten Filaments as Functions of Temperatures, "General Electric Review, pp. 310-319, June 1927; pp. 354-361, July 1927; and pp. 408-412, Aug. 1927
9. Haller, C.E.,"Filament and Heater Characteristics," Electronics, pp. 126-130, July 1944
10. Roberts, N.H., "The Diode as a Half-Wave, Full-Wave, and Voltage-Doubling Rectifier," Wireless Eng. (Brit.), Vol. 13, pp. 351-362, 423-470, July 1936
11. Frommer, J. C., "Operating Data and Allowable Ratings of Vacuum-Tube Rectifiers," Proc. IRE, Vol. 29, No. 9, pp. 481-485, Sept. 1951
12. Aldous, W. H., "The Characteristics of Thermionic Rectifiers, "Wireless Eng. (Brit.), Vol. 13, pp. 576-580, Nov. 1936
13. Kauzmann, A.P., "Determination of Current and Dissipation Values in High-Vacuum Rectifier Tubes," RCA Review, Vol. 8, No. 1, pp. 82-97, March 1947
14. Waidelich, D. L., "Full-Wave Voltage-Doubling Circuits," Proc. IRE, Vol. 29, No. 10, pp. 554-559, Oct. 1941
15. Waidelich, D. L., "The Half-Wave Voltage-Doubling Rectifier," Proc. IRE, Vol. 30, No. 12, pp. 535 ff, Dec. 1942
16. Waidelich, D. L., and C. L.Schackelford,"Characteristics of Voltage Multiplying Circuits," Proc. IRE, Vol. 32, No. 8, pp. 470-476, Aug. 1944
17. Waidelich, D. L., and H. A. K. Taskin, "Analysis of the Voltage Tripling and Quadrupling Circuits," Proc. IRE, Vol. 33, No. 7, pp. 449-457, July 1945
18. Schade, O.H., "Radio-Frequency Operated High-Voltage Supplies for Cathode-Ray Tubes," Proc. IRE, Vol. 31, No. 4, pp. 158-163, April 1943
19. Mautner, R. S., and O. H. Schade, "Television High Voltage RF Supplies," RCA Review, Vol. 8, No. 1, pp. 43-81, March 1947
20. Schade, O.H., "Characteristics of High-Efficiency and High-Voltage Supply Systems for Kinescopes," RCA Review, Vol. 11, No. 1, pp. 5-37, March 1950
21. Deckert, J., "Tube Type 1B3-GT, An Engineering Challenge," RCA Engineer, Vol. 5, No.5, pp. 42-45, Feb.-March 1960
22. Gallup, J., "Glass in Receiving Tubes," RCA Engineer, Vol. 5, No. 6, pp. 38-39, April-May 1960
23. Schade, O. H., "Magnetic-Deflection Circuits for Cathode-Ray Tubes," RCA Review, Vol. 8, No. 3, pp. 506-538, Sept. 1947
24. Murakami, T., "Ringing in Horizontal-Deflection and High-Voltage Television Circuits," RCA Review, Vol. 21, No. 1, pp. 17-44, March 1960
25. Angel, K.W., "Determine Plate Dissipation from Current and Voltage Waveforms," Electronic Design, Vol. 8, No. 11, pp. 50-53, May 1960 (In this article, the plate dissipations of the horizontal-deflection output tubes and of the damper tubes in a typical television circuit are computed from current and voltage waveshape measurements.)

Small-Signal Radio-Frequency Amplifier Tubes

W. A. Harris

Harrison

The design and evaluation of small-signal rf amplifier tubes used in the first stages of television receivers, communications equipment, and similar amplifying apparatus are considered in this article. Noise is a major consideration because the noise produced in tubes and associated circuits determines the extent to which such tubes and circuits are useful in handling small signals. A discussion of the theory of linear amplifiers is also included because the theory provides a means for relating laboratory measurements on tubes to the actual performance obtained from field equipment. In addition, consideration is given to the performance of first-stage tubes functioning in the presence of large signals because many tube applications require good performance over a wide range of input signals.

TELEVISION INPUT-STAGE REQUIREMENTS

A typical application for a small-signal amplifier tube is in the first stage of a very-high-frequency (vhf) television receiver. This type of application imposes the following operating conditions:

1. Operating frequencies are the channel frequencies assigned for television service within the range of 54 to 216 megacycles.

2. Bandwidth for the various stages is determined by the specifications of the television signal. For good detail in the picture, the bandwidth required for the video signal must be 2 to 2.5 megacycles. The bandwidth required for the rf signal must be 6 to 10 megacycles to pass both the picture and the sound carriers and allow some margin for misadjustment and mistuning of the rf circuits.

3. A receiver located near a transmitter may have signals as large as several tenths of a volt developed at the grid of the first rf tube. Under such conditions, the gain of the first tube must be reduced to prevent overloading the second tube.

4. As the distance between the receiver and the transmitter is increased, the receiver signal eventually becomes so small that the picture is obscured by noise. The signal level at which noise overrides the picture is largely determined by the noise factor of the first tube. The smallest usable signals at the grid of the first rf tube are less than 100 microvolts and may be as low as 20 microvolts.

5. For small signals, the gain of the first rf tube must be high enough so that the total noise output of the second rf tube is essentially the amplified noise from the first rf tube. A power gain of 10 or more is indicated by this requirement; however, a higher gain is generally provided because a first rf stage with higher gain permits the use of an intermediate-frequency (if) amplifier of lower gain.

These requirements may be summarized as follows: The first rf tube in a good-quality vhf TV receiver should provide a power gain of at least 10, preferably more, in the frequency range of 50 to 220 megacycles, with a tuner bandwidth of about 10 megacycles. The noise factor should be made as low as possible, subject probably to cost considerations. Input signals up to several tenths of a volt must be handled without excessive distortion if the gain is reduced to unity or a value less than unity. Television applications are challenging because they combine the requirements of low noise, operation with large signals, and operation at high frequencies and substantial bandwidths. Requirements for first-stage tubes in other applications can be similarly analyzed.

MEASUREMENTS OF OPERATING CHARACTERISTICS

Tube design begins with the establishment of design objectives and methods of measuring tube performance. The obvious way of evaluating performance is to test tubes directly in the equipment in which they are used, under simulated field conditions. Indeed, in the early days of broadcasting, this was usually the only method available. Although it is still often necessary to perform such tests, this method of testing is expensive and presents difficulties in obtaining quantitative results.

As an alternative method, tube characteristics which can be conveniently measured at low frequencies are related to actual performance results by factors obtained through correlation tests and theoretical studies. Typical characteristics for such measurements are plate current under chosen operating conditions, transconductance, and capacitance. After a tube design has been standardized for manufacture, the quality and uniformity of the tube can usually be maintained by tests of this type, which can be supplemented by occasional sampling tests more directly related to the tube application. Even the high-frequency noise in a tube can be predicted with fair accuracy from measurements made

at low frequencies of current, transconductance, and capacitance.

When a new tube design is under consideration, or when it is necessary to determine the reasons for differences between tubes of the same type, quantitative measurements of pertinent characteristics in the frequency ranges of interest become necessary. Such tests require careful evaluation, but otherwise present no serious difficulties with the variety and quality of apparatus now available for high-frequency measurements.

A tube under test must be mounted suitably so that dc voltages and signals can be supplied to it. With conventional tube designs, the tube is placed in a socket and supplied with dc power through filters to prevent loss of signal power through the dc leads. Test results obtained in this manner apply to the assembly of tube, socket, and filters rather than to the tube alone. The filters are usually simple; bypass capacitors at heater, cathode, and screen-grid leads backed up by rf chokes serve as low-pass filters for these elements; blocking capacitors in the control-grid and plate leads serve as high-pass filters. Resistors can often be used for filter purposes. However, for high-frequency testing and equipment design, it must be remembered that the leads to capacitors have inductance and that rf chokes and resistors have capacitance between their terminals. Test sets should be designed to keep these undesired factors as small as possible; even then it is sometimes necessary to make corrections for estimated or measured residual impedances.

Measurements which can be made most readily on a high-frequency assembly such as that just described are the terminal impedances (or admittances, the reciprocal values) and the power gain. Noise factor can be measured if an amplifier of sufficient gain to bring the noise power output from the unit tested up to a measurable level is connected to the output terminals of the test assembly.

DEFINITIONS AND THEORY

Before proceeding with the discussion of test methods, some of the quantities measured at high frequencies must be defined and the theoretical basis for the measurements discussed.

Transducer: a device actuated by power from one system and supplying power to a second system. Tubes with operating voltages applied, attenuators, loudspeakers, and kinescopes are all forms of transducers.

An operating tube and its associated circuits can be considered a linear transducer when the signals handled by the tube are small. Linearity implies exact proportionality between input and output signals. Although no active transducer is exactly linear, it may be considered linear in the range where there is satisfactory agreement between measured performance and performance predicted from calculations based on the assumption that the transducer is linear.

Two-port transducer (or two-port, or a four-pole, or a two-terminal-pair transducer): a transducer with only two places for the introduction and removal of signals. The test assembly previously discussed is a two-port transducer.

Active two-port: a two-port in which power other than the input-signal power is used in processing the signal. The test assembly discussed is an active two-port when the heater voltage and other tube voltages are applied. When these voltages are removed, it becomes a passive two-port. Attenuators, transformers, most loudspeakers, and transmission lines are passive two-ports.

Available power from a port: the power which would be supplied from the port to a termination adjusted to receive maximum power, i. e., to a "matching" load. The available power from a signal generator is measured by connecting a power meter, which may use a thermistor of perhaps a thermocouple, to the generator terminals through a transformer to match the impedance of the power meter to the impedance of the generator.

Impedance or admittance at a port: These terms can be defined when the current and voltage at the port can be determined. When the port consists of a pair of terminals or a transmission-line coupling, the idea of impedance or admittance presents no difficulty. When the port is a waveguide junction or a prescribed closed surface in space around an antenna, it is still easy to discuss available power, but voltage, current, and admittance may be difficult to define. In the cases discussed in this article, impedance or admittance is easily defined and measured.

Noise factor: the ratio of the total noise power available or delivered from the output port of a transducer to that portion of the output power which comes from the thermal-agitation noise developed by the input termination, when the input termination is passive and at a standard temperature of 290 K.

When the noise factor of a transducer has been determined for a given input termination, it is possible to determine from this value the noise performance to be expected from the same transducer connected to an antenna or to any other termination having the same impedance.

Transducer gain of a two-port: the ratio of the power delivered to an output termination to the available power from the input termination.

Available gain: the ratio of the available power from the output port to the available power from the input termination. Available gain is identical to transducer gain when the output termination is matched to the output port.

Measurements of a Two-Port

The performance of a linear two-port can be determined completely by a relatively small number of measurements made at the ports. Consider a two-port consisting of an assembly comprising a tube, a socket, and filters to allow introduction of dc power and heater

power, and input and output terminals for the introduction and removal of signal power. Because measurements are to be made at high radio frequencies, it is convenient to use coaxial connectors as the input and output terminals. The properties of this kind of a two-port can be expressed in terms of the admittances at the terminals; both transducer gain and noise factor are functions of the terminal admittances and the frequency. The following set of parameters and their symbols can be used to describe the performance of a two-port:

Short-circuit input conductance - G_{11}
Short-circuit input susceptance - B_{11}
Short-circuit output conductance - G_{22}
Short-circuit output susceptance - B_{22}
Short-circuit forward admittance magnitude - $|Y_{21}|$
Short-circuit forward admittance phase angle - ϕ_{21}
Short-circuit feedback admittance magnitude - $|Y_{12}|$
Short-circuit feedback admittance phase angle - ϕ_{12}

The symbol ϕ without subscript refers to the phase angle of the product of the short-circuit forward and feedback admittances.

In the diagram shown in Fig. 1, the box (Y) can represent any linear two-port. The symbol (Y) represents the matrix $\begin{pmatrix} Y_{11} & -Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}$. Currents I_1 and I_2 are shown entering the input and output ports, respectively, and voltages V_1 and V_2 are represented across the input and output ports. A terminating admittance Y_0 is shown connected to the input port and a terminating admittance Y_3 is shown connected to the output port.

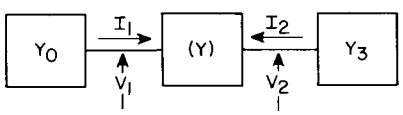


Figure 1. Linear Two-Port and Terminations

The following equations show the relationship of these currents and voltages:

$$I_1 = Y_{11}V_1 - Y_{12}V_2 \quad (1)$$

$$I_2 = -Y_{21}V_1 + Y_{22}V_2 \quad (2)$$

These equations define the short-circuit admittances. When the output port is short-circuited, V_2 is zero; the admittance Y_{11} is the quotient I_1/V_1 and the admittance Y_{21} is the negative of the quotient I_2/V_1. Similar considerations apply to the other two admittances Y_{12} and Y_{22}.

The short-circuit admittances are represented by complex numbers. The real and imaginary parts of the short-circuit input and output admittances are the conductances G_{11} and G_{22} and the susceptances B_{11} and B_{22}. The short-circuit forward and feedback admittances can be divided in the same manner, but in the development of the equations for gain it is more convenient to use the magnitudes of these admittances $|Y_{21}|$ and $|Y_{12}|$, and the phase angles ϕ_{21} and ϕ_{12}. The phase angle ϕ of the product of the forward and feedback admittances is the sum of the phase angles ϕ_{21} and ϕ_{12}.

Fig. 1 shows that the current into the output termination Y_3 is the negative of the current I_2 shown directed into the two-port (Y). Consequently,

$$I_2 = -Y_3V_2 \quad (3)$$

The power supplied to the output termination Y_3 is the product of the conductance component G_3 of the admittance Y_3 and the mean-square voltage $\overline{V_2^2}$ across Y_3. The general symbols for mean-square voltage and current are $\overline{V^2}$ and $\overline{I^2}$. When the symbols for voltage and current (V_1, V_2, I_1, I_2) represent the root-mean-square amplitudes of sinusoidal voltages with appropriate phase angles, the corresponding mean-square quantities can be represented as squares of magnitudes, i.e., as $|V_1|^2$, $|V_2|^2$, $|I_1|^2$, and $|I_2|^2$. The square of the magnitude of a complex number can be found by multiplying the number by its conjugate, thus:

$$|V_2|^2 = V_2V_2^*$$

The asterisk indicates the conjugate of a complex number, which is the number with its imaginary part reversed in sign.

Using the convention that current and voltage magnitudes are root-mean-square values, the power output P_2 from (Y) in Fig. 1 is expressed as follows:

$$P_2 = |V_2|^2 G_3 \quad (4)$$

The input power P_1 is similarly expressed as:

$$P_1 = |V_1|^2 G_1 \quad (5)$$

where G_1 is the real part of the input admittance of the two-port (Y) with the output termination Y_3 connected. The value of G_1 may be found as follows:

From Eqs. (2) and (3),

$$-Y_3 V_2 = -Y_{21} V_1 + Y_{22} V_2$$

then,

$$V_2 = V_1 Y_{21}/(Y_{22} + Y_3) \quad (6)$$

Substituting this expression for V_2 in Eq. (1) gives

$$I_1 = V_1 \left[Y_{11} - Y_{12}Y_{21}/(Y_{22} + Y_3)\right]$$

However, the input admittance Y_1 is the quotient of the input current I_1 divided by the input voltage V_1; therefore,

$$Y_1 = Y_{11} - Y_{12}Y_{21}/(Y_{22} + Y_3) \quad (7)$$

Consequently, the conductance G_1 can be represented as the real part of the right hand term of Eq. (7):

$$G_1 = \mathrm{Re}\left[Y_{11} - Y_{12}Y_{21}/(Y_{22} + Y_3)\right] \quad (7a)$$

Then, from Eqs. (5) and (7a):

$$P_1 = |V_1|^2 \, \mathrm{Re}\left[Y_{11} - Y_{12}Y_{21}/(Y_{22} + Y_3)\right] \quad (8)$$

Squaring the magnitudes in Eq. (6) gives:

$$|V_2|^2 = |V_1|^2 \, |Y_{21}|^2 / |Y_{22} + Y_3|^2 \quad (9)$$

Substituting this equation in Eq. (4) gives:

$$P_2 = |V_1|^2 | \, Y_{21}|^2 \; G_3 / |Y_{22} + Y_3|^2 \quad (10)$$

The power gain PG is obtained by dividing Eq. (10) by Eq. (8) as follows:

$$PG = |Y_{21}|^2 \, G_3 / |Y_{22} + Y_3|^2 \, \mathrm{Re}\left[Y_{11} - Y_{12}Y_{21} / (Y_{22} + Y_3)\right] \quad (11)$$

For a stable system, the power gain as given by Eq. (11) must be positive. When the output termination admittance has a positive real part (as would be the case for any passive load connected to the two-port), the numerator in Eq. (11) is positive; therefore, the requirement for stability is that the input admittance Y_1 [See Eq. (7)] must have a positive real part.

In Fig. 2, SG represents a signal generator having a known output impedance and a known open-circuit output voltage. T_0 represents a loss-free transformer which can transform the output impedance of the signal generator to any desired value of admittance, corresponding to the input termination admittance Y_0 of Fig. 1. PM represents a calibrated power output meter with a known input impedance; T_3 represents a loss-free transformer which can transform the input impedance of the power output meter to any desired value of admittance, corresponding to the output termination Y_3 of Fig. 1. The available power from a signal generator is the power supplied by the generator to a terminating impedance which is the conjugate of the internal impedance of the generator. If the root-mean-square open-circuit voltage of the generator is V_g and the resistive component of its output impedance is R_g, the available power AP is expressed as follows:

$$AP = V_g^2 / 4R_g$$

When a loss-free transformer, such as T_0 in Fig. 2, is connected to the generator, the available power at the output port of T_0 is the same as the available power at the output port of the generator. When the output admittance Y_0 of the transformer T_0 is made equal to the complex conjugate of the input admittance Y_1 of the two-port (Y), the input power P_1 is equal to the available power of the generator. The output power P_3 is measured directly by the power output meter PM. Consequently, the power gain PG is the ratio of the power indicated by the output meter PM to the available power of the generator AP. The transformer T_0 of Fig. 2 can be realized physically in many ways. Some of these ways are suggested in Fig. 3.

For frequencies of the order of 30 to 300 megacycles per second, lumped-constant networks of various configurations are generally used. For frequencies from 300 to 3000 megacycles, the use of transmission lines and tuning stubs is more satisfactory. At still higher frequencies, waveguide components are used. A general requirement to be met is that there be two degrees of freedom for adjustment, so that any desired combination of real and imaginary components of the output admittance can be obtained.

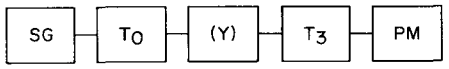


Figure 2. Measurement of Power Gain

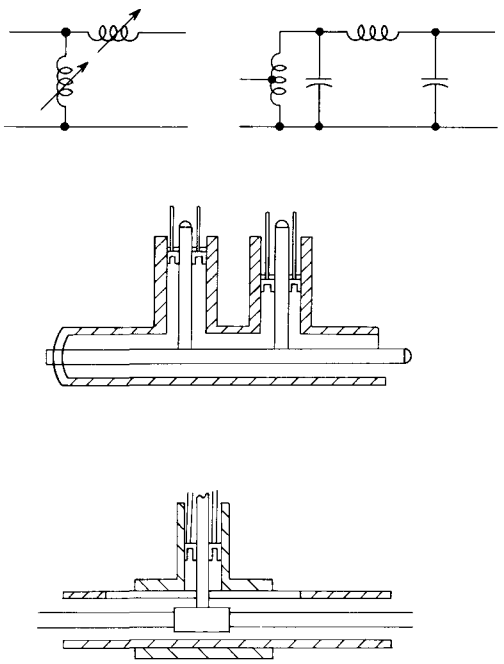


Figure 3. Admittance Transformers

Test equipment such as the RX meter or the Q meter (both made by Boonton Radio Corporation), or the General Radio Admittance Meter, can be used to measure the conductances and susceptances at the terminals of a two-port at frequencies up to a few hundred megacycles. Standing-wave-ratio measurements made with slotted lines can be used at higher frequencies; then, the conductances and susceptances can be found by calculation or by graphical methods, e.g., with the Smith Chart.

The foregoing considerations indicate that the power gain and the terminal conductance and terminal-admittance terms represented in Eq. (11) can be measured. The magnitude of the short-circuit forward admittance $|Y_{21}|$ and the short-circuit feedback admittance $|Y_{12}|$, and also the phase angle ϕ of the product $Y_{12}Y_{21}$ can be found by making several (at least three) power gain measurements for different values of the output termination admittance components, G_3 and B_3, writing the results in the form of Eq. (11), and solving for the desired quantities. Some additional properties of a linear

two-port can be determined by examining Eq. (7) for the input admittance Y_1.

$$Y_1 = Y_{11} - Y_{12}Y_{21}/(Y_{22} + Y_3) \quad (7)$$

When the exponential form for the representation of complex numbers is used, the product $Y_{12}Y_{21}$ can be represented as follows:

$$Y_{12}Y_{21} = |Y_{12}Y_{21}|\ e^{j\phi} \quad (12)$$

If the phase angle for the sum $(Y_{22} + Y_3)$ is designated as the angle θ this sum can be represented by

$$Y_{22}+Y_3 = |Y_{22} + Y_3|\ e^{j\theta}$$

But, because $e^{j\theta} = \cos\theta + j \sin\theta$, the real and imaginary components of this sum are

$$G_{22}+G_3 = |Y_{22} + Y_3| \cos\theta$$

$$B_{22}+B_3 = |Y_{22} + Y_3| \sin\theta$$

Therefore,

$$\theta = \tan^{-1}(B_{22} + B_3)/(G_{22} + G_3) \quad (13)$$

and

$$|Y_{22} + Y_3| = (G_{22} + G_3)/\cos\theta$$

and therefore,

$$Y_{22}+Y_3 = (G_{22} + G_3)\, e^{j\theta}/\cos\theta \quad (14)$$

Substituting Eqs. (12) and (14) in Eq. (7) gives:

$$Y_1 = Y_{11} \frac{|Y_{12}Y_{21}|}{G_{22}+G_3} e^{j\phi} \cos\theta\, e^{-j\theta}$$

But the cosine function can be represented by

$$\cos\theta = \frac{1}{2}(e^{j\theta} + e^{-j\theta})$$

Therefore,

$$Y_1 = Y_{11} - \frac{1}{2}\frac{|Y_{12}Y_{21}|}{G_{22}+G_3} e^{j\phi}\ (1 + e^{-j2\theta}) \quad (15)$$

Eq. (13) shows that when the real component G_3 of the output termination admittance is held constant, the phase angle θ of the sum $(Y_{22} + Y_3)$ of the short-circuit output admittance and the output termination admittance is a function of the imaginary component B_3 only.

The significance of the terms of Eq. (15) is demonstrated by plotting the terms of this equation in the complex plane, as shown in Fig. 4. Eq. (15) and Fig. 4 show that when the output termination conductance G_3 is held constant and the angle θ is varied by varying the output termination susceptance B_3, the locus of the input admittance Y_1 plotted in the complex plane is a circle having a radius which can be expressed as follows:

$$\text{Rad} = \frac{1}{2}\frac{|Y_{12}Y_{21}|}{G_{22}+G_3}$$

The center is located at

$$C = Y_{11} - \frac{1}{2}\frac{|Y_{12}Y_{21}|}{G_{22}+G_3} e^{j\phi}$$

and the angle θ is measured from a point P on the circle given by

$$P = Y_{11} - \frac{|Y_{12}Y_{21}|}{G_{22}+G_3} e^{j\phi}$$

When a family of input admittance curves is plotted, using different values of the output termination conductance G_3 to obtain the different plots, the resulting circles all pass through the point representing Y_{11}, and the centers of all the circles are on the same straight line. Such a family of circles is shown in Fig. 5.

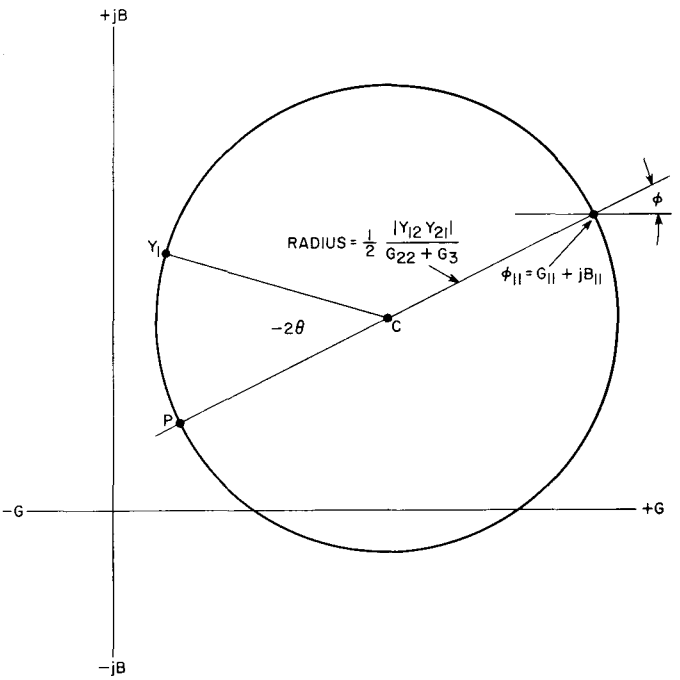


Figure 4. Locus of Input Admittance Equation

The condition that the real part G_1 of the input admittance Y_1 be positive means that (1) the short-circuit input admittance Y_{11} must have a positive real part and (2) the smallest value of the output termination conductance G_3 to be used must produce a circle which does not cross into the region of negative conductance; or (3) the output susceptance B_3 must be restricted to values giving values of Y_1 with positive real parts. A two-port is unconditionally stable when no part of the circle representing Y_1 for the case $(G_3 = 0)$ falls in the region of negative conductance.

In a practical sense, adjustment of B_3 can be thought of as the tuning of the output circuit of an amplifier and adjustment of G_3 can be thought of as adjustment of coupling to a load or, in some cases, as the choosing of a resistor value to provide a load for an amplifier.

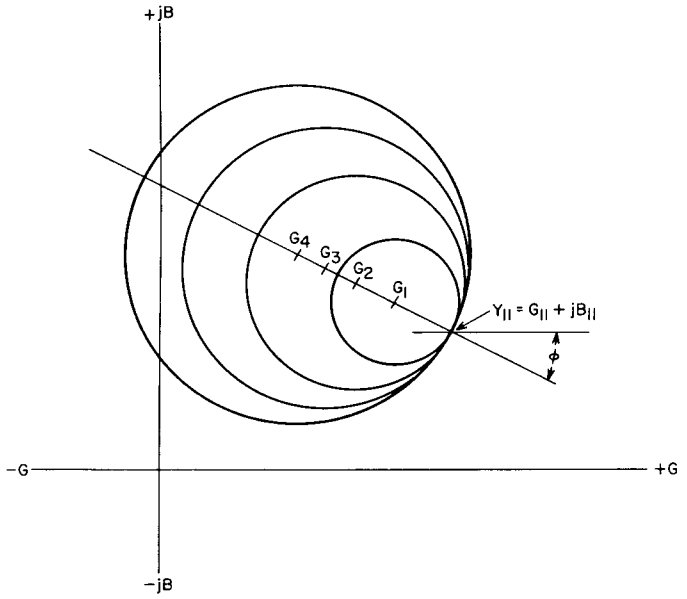


Figure 5. Admittance Loci for Varied Conductance

Three-Terminal Transducer

A triode amplifier may be used in any of three different basic amplifier circuits which are designated as grounded-cathode, grounded-grid, or grounded-plate circuits. In each case the grounded electrode is connected to the low-potential point in the system under consideration through an impedance which is low at the frequency of the signal to be amplified.

The grounded-plate circuit is often referred to as a cathode-follower circuit. The grounded-grid circuit is sometimes described as a grid-separation circuit. The grounded-cathode circuit is the earliest of the triode amplifier circuits and is the circuit most often implied when an amplifier is discussed without reference to the way it is connected. Tetrodes and pentodes are most frequently used in grounded-cathode circuits. But it is also possible to use these types in grounded-grid circuits.

The triode can be considered as an example of a three-terminal transducer which can be used to advantage for certain applications with any one of the three terminals common to the input and output circuits. A general theory for three-terminal transducers follows.

In Fig. 6 the three terminals are designated A, B, and C; the signal-frequency currents into these terminals are I_A, I_B, and I_C; and the voltages at the terminals are V_A, V_B, and V_C.

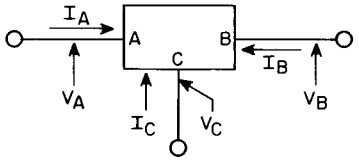


Figure 6. Three-Terminal Transducer

Nine admittance coefficients can be defined by the following three equations:

$$I_A = V_A Y_{AA} - V_B Y_{AB} - V_C Y_{AC} \quad (16)$$

$$I_B = -V_A Y_{BA} + V_B Y_{BB} - V_C Y_{BC} \quad (17)$$

$$I_C = -V_A Y_{CA} - V_B Y_{CB} + V_C Y_{CC} \quad (18)$$

These equations are not independent. First, in the case when all three voltages are made equal, that is $V_B = V_C = V_A$, the currents are all zero because there is no voltage between any pair of terminals, and the equations become:

$$0 = V_A(Y_{AA} - Y_{AB} - Y_{AC})$$

$$0 = V_A(-Y_{BA} + Y_{BB} - Y_{BC})$$

$$0 = V_A(-Y_{CA} - Y_{CB} + Y_{CC})$$

Therefore:

$$Y_{AA} = Y_{AB} + Y_{AC} \quad (19)$$

$$Y_{BB} = Y_{BA} + Y_{BC} \quad (20)$$

$$Y_{CC} = Y_{CA} + Y_{CB} \quad (21)$$

Adding Eqs. (16), (17), and (18) gives:

$$I_1 + I_2 + I_3 = V_A(Y_{AA} - Y_{BA} - Y_{CA}) + V_B(Y_{BB} - Y_{AB} - Y_{CB}) + V_C(Y_{CC} - Y_{AC} - Y_{BC})$$

But the sum of the currents must be zero.

Therefore,

$$V_A(Y_{AA} - Y_{BA} - Y_{CA}) + V_B(Y_{BB} - Y_{AB} - Y_{CB}) + V_C(Y_{CC} - Y_{AC} - Y_{BC}) = 0$$

for any values that can be given to the voltages V_A, V_B, and V_C.

Consequently,

$$Y_{AA} = Y_{BA} + Y_{CA} \quad (22)$$

$$Y_{BB} = Y_{AB} + Y_{CB} \quad (23)$$

$$Y_{CC} = Y_{AC} + Y_{BC} \quad (24)$$

The six equations, (19) to (24) inclusive, are not independent because adding the three equations of either group gives the result:

$$Y_{AA} + Y_{BB} + Y_{CC} = Y_{AB} + Y_{AC} + Y_{BA} + Y_{BC} + Y_{CA} + Y_{CB} \quad (25)$$

Consequently, if any five of the equations are given, the sixth equation can be found by adding these five and

subtracting the sum from Eq. (25) doubled. However, any five of these equations may be chosen and used to reduce the nine admittance coefficients in Eqs. (16), (17), and (18) to quantities involving the sums and differences of only four coefficients.

For example;

$$\text{From Eq. (19)} \quad Y_{AC} = Y_{AA} - Y_{AB} \tag{26}$$

$$\text{From Eq. (20)} \quad Y_{BC} = Y_{BB} - Y_{BA} \tag{27}$$

$$\text{From Eq. (22)} \quad Y_{CA} = Y_{AA} - Y_{BA} \tag{28}$$

$$\text{From Eq. (23)} \quad Y_{CB} = Y_{BB} - Y_{AB} \tag{29}$$

Adding Eqs. (28) and (29) gives:

$$Y_{CA} + Y_{CB} = Y_{AA} + Y_{BB} - Y_{BA} - Y_{AB} \tag{30}$$

Substituting Eq. (30) in Eq. (21) gives:

$$Y_{CC} = Y_{AA} + Y_{BB} - Y_{BA} - Y_{AB} \tag{31}$$

When V_C is zero, then from Eqs. (16) and (17):

$$I_{A(C)} = V_A Y_{AA} - V_B Y_{AB} \tag{32}$$

$$I_{B(C)} = -V_A Y_{BA} + V_B Y_{BB} \tag{33}$$

These equations are identical in form with the two-port equations, Eqs. (1) and (2). A convenient change in notation gives:

$$I_{1(C)} = V_{1(C)}\, Y_{11(C)} - V_{2(C)}\, Y_{12(C)} \tag{34}$$

$$I_{2(C)} = V_{2(C)}\, Y_{21(C)} - V_{2(C)}\, Y_{22(C)} \tag{35}$$

The identities

$$Y_{11(C)} \equiv Y_{AA} \tag{36}$$

$$Y_{12(C)} \equiv Y_{AB} \tag{37}$$

$$Y_{21(C)} \equiv Y_{BA} \tag{38}$$

$$Y_{22(C)} \equiv Y_{BB} \tag{39}$$

are thus established for the case that terminal C of the three-terminal transducer is grounded.

The case in which V_A is zero, GA is the input port, and BA is the output port, leads to the identities:

$$Y_{11(A)} \equiv Y_{CC} \tag{40}$$

$$Y_{12(A)} \equiv Y_{CB} \tag{41}$$

$$Y_{21(A)} \equiv Y_{BC} \tag{42}$$

$$Y_{22(A)} \equiv Y_{BB} \tag{43}$$

Substituting Eqs. (31), (29), and (27) in Eqs. (40), (41), and (42) gives:

$$Y_{11(A)} = Y_{AA} + Y_{BB} - Y_{BA} - Y_{AB} \tag{44}$$

$$Y_{12(A)} = Y_{BB} - Y_{AB} \tag{45}$$

$$Y_{21(A)} = Y_{BB} - Y_{BA} \tag{46}$$

$$Y_{22(A)} = Y_{BB} \tag{47}$$

Triode Admittances

Fig. 7 shows a triode connected in a grounded-grid circuit. Signal-voltage generators are used to apply signal-frequency voltages V_K and V_P to the cathode and plate terminals. Signal-frequency currents I_K and I_P, into the cathode and the plate, respectively, are read with meters I_K and I_P.

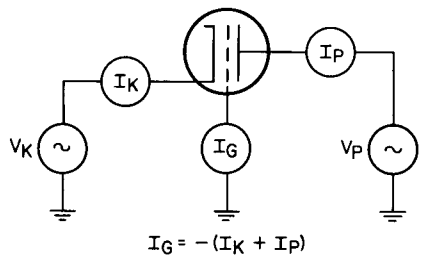


Figure 7. Grounded-Grid Triode

The signal-frequency currents indicated in the circuit of Fig. 7 are determined by the signal-frequency voltages V_K and V_P, the capacitances between the electrodes, and the signal-frequency component of the electron current in the tube. The signal-frequency component of electron current to the plate at low frequencies is determined by the transconductance G_m and the plate conductance G_p (the reciprocal of the plate resistance R_p). When the signal frequency is so low that the currents through the tube capacitances can be neglected, the short-circuit admittances are:

$$Y_{11} = 0$$

$$Y_{12} = 0$$

$$Y_{21} = G_m + G_p$$

$$Y_{22} = G_p$$

However, the plate conductance G_p of a triode is the transconductance G_m, divided by the amplification factor μ; therefore,

$$G_p = 1/\mu$$

$$Y_{21} = G_m\,(1 + 1/\mu)$$

$$\text{and } Y_{22} = Y_{21}/(1 + \mu)$$

When the signal frequency is high enough that capacitances must be taken into account, the effect of the phase delay in the signal-frequency component of the electron current must also be considered. This phase delay is caused by and is a function of the time required

for electrons to travel from cathode to grid and from grid to plate. Some of the effects of phase delay include: (1) A decrease in the capacitance of a diode carrying current, as compared with the capacitance of the same diode when current is absent. (2) An increase in the grid-to-cathode capacitance of a triode, tetrode, or pentode (operated in a grounded-cathode circuit) carrying current, as compared with the grid-to-cathode capacitance of the same tube in the absence of current. (3) A conductance component of the input admittance of a triode, tetrode, or pentode (operated in a grounded-cathode circuit) increasing approximately with the square of the signal frequency. (4) The phase delay itself, observable as a time delay for a signal current waveform at the plate as compared with the input voltage waveform at the grid or cathode.

Because the capacitance changes have often been treated as phenomena in themselves without reference to their relation to transit time and because small phase delays are not easy to identify in signal-handling systems, there has been a tendency to treat transit-time effects as if they were of no consequence at signal frequencies below those at which the input conductance becomes prominent. Neglecting the phase delay in the transconductance leads to confusion and wrong answers when the three-terminal analysis discussed in this article is applied to triodes.

A suitable procedure for determining the short-circuit admittance parameters for a triode follows.

First, consider the diode consisting of the cathode and the control-grid surface. When there is no emission of electrons, a signal-frequency voltage V across the diode produces a current Vj ωC, in phase quadrature with the voltage, because of the diode capacitance. When there is emission of electrons and the dc voltage is such as to establish a space-charge-limited current through the diode, there is also a signal-frequency component of current in phase with the voltage. At the same time the quadrature component is reduced because of the phase delay caused by the inertia of the electrons. An approximate equation for the admittance of a space-charge-limited diode is: $Y_D = G_D + K_D j \omega C$. The factor K_D is approximately 0.6 when the dc diode voltage is approximately one volt or more.

In the case of a triode with a negative dc voltage on the grid, the part of the current represented by motion of electrons passes through the grid, and there is some additional delay in phase in the passage of the current from the grid to the plate. One requirement that must be satisfied for the triode (and for any three-terminal transducer) is that the sum of the three currents flowing into the cathode, grid, and plate terminals must be zero. Fig. 8 shows the relations between the signal-frequency cathode voltage and the signal-frequency currents into the cathode, plate, and grid when a signal-frequency voltage is applied to the cathode and the signal-frequency voltages at the grid and at the plate are zero.

A vector V_K drawn horizontally to the right, represents the signal-frequency cathode voltage. The vector representing current into the cathode I_K is drawn with a positive phase angle relative to the vector representing the cathode voltage.

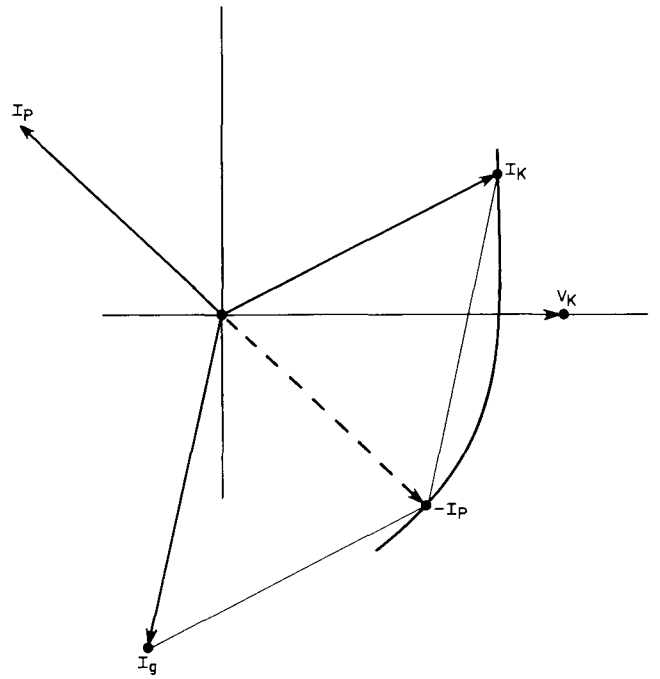


Figure 8. Triode Currents with Voltage to Cathode

The vector representing the negative of the signal-frequency plate current (that is, current out of the plate terminal) is represented as having the same magnitude as the projection of the cathode-current vector I_K on the cathode-voltage vector V_K, but rotated in phase in the negative direction from V_K. This vector is shown as a dotted line and is labeled $-I_p$. The vector representing the plate current I_p into the plate terminal is drawn as a solid line of the same length, in the opposite direction.

The requirement that the sum of the currents into the grid, plate, and cathode must be zero is satisfied by drawing the vector I_g representing current into the grid terminal with length equal to and direction parallel to the line completing the triangle formed by the vectors representing the negative of the plate current $-I_p$ and the cathode current I_K.

Fig. 9 shows the relation between the signal-frequency plate voltage and the signal-frequency currents into the plate, cathode, and grid when a signal-frequency voltage is applied to the plate and the signal-frequency voltages at the cathode and at the grid are zero. In this case, the signal-frequency current into the plate terminal has two components. The first component I_{p1}, in Fig. 9, may be considered as resulting from an effective signal voltage at the grid proportional to the plate-to-grid voltage V_p. The phase angle between the vector representing I_{p1} and the vector representing V_p in Fig. 9 is the same as the phase angle between the vector representing $-I_p$ and the vector representing V_K in Fig. 8, because in both cases the plate current results from a voltage between cathode and grid and because this voltage is in phase with V_p in the case of

Fig. 9. Because the second component, I_{p2} in Fig. 9, is the current through the grid-to-plate capacitance (C_{gp}), the vector representing I_{p2} is drawn at right angles to the vector representing V_p, and is rotated in the positive direction with respect to V_p. The vector representing the total signal-frequency current into the plate terminal, labeled I_p in Fig. 9, is found by adding the vectors representing the components I_{p1} and I_{p2}.

The vector representing the negative of the cathode current (that is, current out of the cathode terminal) is shown as a dotted line and is labeled $-I_K$. This vector is advanced in phase with respect to V_p by the same amount that the vector I_K in Fig. 8 is advanced with respect to V_K because in both cases the cathode current results from a voltage between grid and cathode and because this voltage is in phase with V_p in the case of Fig. 9. The vector I_K representing current into the cathode terminal is drawn in the opposite direction to $-I_K$ and has the same length. The vector I_g representing signal-frequency current into the grid terminal is drawn parallel to the line completing the triangle formed by the vectors representing the plate current I_p and the negative of the cathode current $-I_K$.

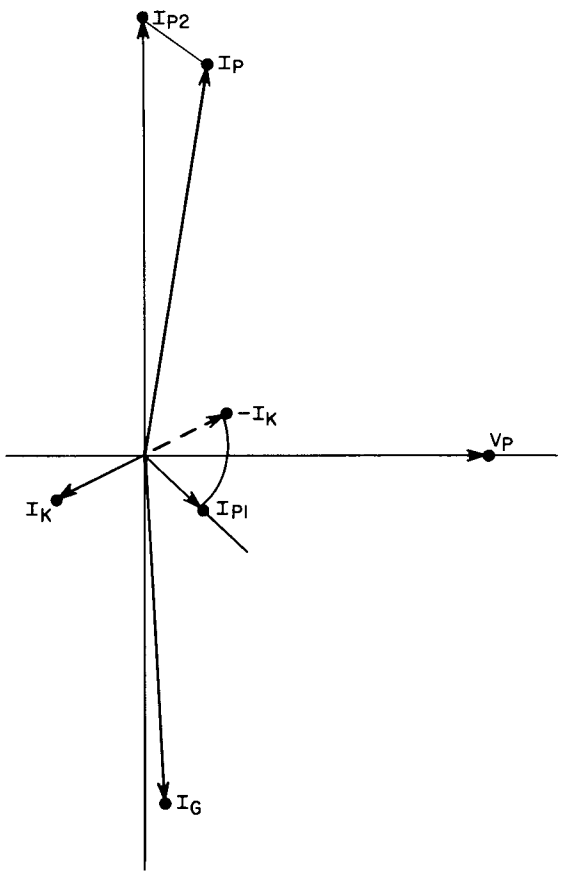


Figure 9. Triode Currents with Voltage to Plate

The following equations describe the vectors of Fig. 8:

$$I_k = V_K (G_m + G_p + j\omega C_D) \tag{48}$$

$$I_p = -V_K (G_m + G_p)\, e^{-j\omega\tau pk}$$

$$= -V_K(G_m + G_p)\cos\omega\tau_{pk} + jV_K(G_m + G_p)\sin\omega\tau_{pk} \tag{49}$$

$$I_g = -(I_p + I_k)$$

$$= -V_K(G_m + G_p)(1 - \cos\omega\tau_{pk}) - jV_K\left[\omega C_D + (G_m + G_p)\sin\omega\tau_{pk}\right] \tag{50}$$

The following equations describe the vectors of Fig. 9:

$$I_K = -V_p\left(G_p + j\omega\frac{C_D}{1+\mu}\right) \tag{51}$$

$$I_p = V_p\, G_p\, e^{-j\omega\tau pk} + jV_P\omega C_{gp}$$

$$= V_P\, G_p \cos\omega\tau_{pk} + jV_P(-G_p\sin\omega\tau_{pk} + \omega C_{gp}) \tag{52}$$

$$I_g = -(I_P + I_K)$$

$$= V_P G_P(1-\cos\omega\tau_{pk}) - jV_P\left[\omega\left(C_{gp} - \frac{C_D}{1+\mu}\right) + G_p\sin\omega\tau_{pk}\right] \tag{53}$$

The symbols used in Eqs. (48) through (53) are defined as follows:

G_m is the grid-to-plate transconductance, as defined and measured at a low frequency

G_p is the plate conductance, as defined and measured at a low frequency

μ is the amplification factor, as measured under tube operating conditions at a low frequency

τ_{pk} is the phase delay in seconds between current at the plate and voltage between grid and cathode

ω is the angular frequency, in radians per second

C_D is the effective grid-to-cathode capacitance measured under operating conditions with signal-frequency voltage applied to the cathode and with the grid and plate grounded with respect to the signal frequency.

The short-circuit admittance coefficients may be found by inspection of Eqs. (48), (49), (51), and (52) and comparison of these equations with Eqs. (1) and (2). Correspondence with the three-terminal-transducer equations, Eqs. (36) through (39), is also evident. This correspondence may be used to obtain the short-circuit admittances for grounded-grid operation as follows:

$$Y_{11} = G_m + G_p + j\omega C_D = Y_{AA} \tag{54}$$

$$Y_{12} = G_p + j\omega\frac{C_D}{1+\mu} = Y_{AB} \tag{55}$$

$$Y_{21} = (G_m + G_p)\, e^{-j\omega\tau\, pk}$$

$$= (G_m + G_p)\cos\omega\tau_{pk} - j(G_m + G_p)\sin\omega\tau_{pk} = Y_{BA} \tag{56}$$

$$Y_{22} = G_p \cos\omega\tau_{pk} - jG_p \sin\omega\tau_{pk} + j\omega C_{gp} = Y_{BB} \quad (57)$$

The short-circuit admittances for grounded-cathode operation can be found by using Eqs. (44) to (47).

The resulting equations are:

$$Y_{11(K)} = Y_{11} + Y_{22} - Y_{12} - Y_{21}$$
$$= G_m(1 - \cos\omega\tau_{pk}) + j\left[\omega C_D\left(\frac{\mu}{1+\mu}\right) + G_m \sin\omega\tau_{pk} + \omega C_{gp}\right] \quad (58)$$

$$Y_{12(K)} = Y_{22} - Y_{12}$$
$$= G_p(\cos\omega\tau_{pk} - 1) + j\left[\omega C_{gp} - G_p \sin\omega\tau_{pk} - \omega\frac{C_D}{1+\mu}\right] \quad (59)$$

$$Y_{21(K)} = Y_{22} - Y_{21}$$
$$= -G_m e^{-j\omega\tau_{pk}} + j\omega C_{gp}$$
$$= -G_m \cos\omega\tau_{pk} + j\left[\omega C_{gp} + G_m \sin\omega\tau_{pk}\right] \quad (60)$$

$$Y_{22(K)} = Y_{22}$$
$$= G_p \cos\omega\tau_{pk} + j\left[\omega C_{gp} - G_p \sin\omega\tau_{pk}\right] \quad (61)$$

For small values of the phase angle ($\omega\tau_{pk}$),

$$\sin\omega\tau_{pk} \doteq \omega\tau_{pk} \quad (62)$$

$$\cos\omega\tau_{pk} \doteq 1 - \frac{1}{2}(\omega\tau_{pk})^2 \quad (63)$$

The phase delay time τ_{pk} is of the order of 10^{-10} seconds for tubes having dimensions suitable for application as rf amplifiers in television service. The resulting phase angle at 200 megacycles per second is 0.125 radian. Consequently, the approximations from Eqs. (62) and (63) may be used for this frequency. From Eq. (58), the input conductance is $G_{11} = 1/2G_m\omega^2(\tau_{pk})^2$. This equation shows that the input conductance is proportional to the transconductance, the square of the signal frequency, and the square of the phase-delay time, when the approximation indicated by Eq. (63) is allowable. Other methods of demonstrating the existence of input conductance for a triode, tetrode, or pentode used in a grounded-cathode circuit give results which can be identified with the result obtained here.

The above discussion shows that the input-conductance term results directly from the phase delay of the plate current with respect to the grid-to-cathode voltage, and suggests that any added phase delay, such as might be caused by addition of inductance in series with the cathode of a tube, would increase the input conductance. In fact, there is no simple method of discriminating between the input-conductance components caused by phase delay chargeable to electron transit time and input-conductance components caused by inductance in the cathode lead of a tube or inductance in the external cathode connection.

Comparison of Eq. (54) with Eq. (58) shows that the input-conductance term varying with the square of the frequency (in the grounded-cathode configuration) does not appear as part of the input conductance for the grounded-grid configuration. At low frequencies the low value of input conductance obtainable with the grounded-cathode configuration is an important consideration favoring the use of this form of circuit, but at high frequencies this advantage is lost.

The input capacitance for the grounded-cathode circuit can be found by dividing the imaginary part of the short-circuit input admittance given by Eq. (58) by the angular frequency. The result, with the approximation of Eq. (62), is $C_{11(K)} = C_D(\mu/1+\mu) + C_{gp} + G_m\tau_{pk}$. For a transconductance of 10 millimhos and a phase delay time of 10^{-10} seconds, the last term gives the equivalent of a capacitance of 1 micromicrofarad to be added to the other terms.

The principal advantage obtained by the use of the grounded-grid circuit at high frequencies is the reduction of feedback. Eq. (55) shows that the feedback for the grounded-grid circuit is just the short-circuit input admittance divided by the quantity $(1 + \mu)$. Eq. (63) shows that the largest part of the feedback for the grounded-grid case is the susceptance of the grid-to-plate capacitance. Circuits with external feedback adjusted in phase and magnitude to neutralize the effect of the grid-to-plate capacitance are used at vhf television frequencies to obtain tolerable values of feedback along with the lower input conductance values of the grounded-cathode circuit. Tetrodes and pentodes may also be used as rf amplifiers at vhf television frequencies, but these tubes give substantially higher noise factors than triodes.

Measurement of Noise

The noise factor of a receiver or an amplifier is usually determined by connecting a calibrated noise source to the input terminals of the device, observing the power output, and comparing this power output with the power output observed when a passive termination at standard temperature is connected to the input terminals. A temperature-limited diode connected to a resistor of known value is frequently used as the noise source. The diode produces noise because the emission of electrons from its filament is random, i.e., it depends on statistically independent events. The short-circuit mean-square noise current from a temperature-limited diode (i.e., a diode operating with negligible space charge) is given by the equation

$$\overline{i^2} = 2\,eI\Delta f \quad (64)$$

where $\overline{i^2}$ is the mean-square noise current in amperes squared; I is the dc current in amperes; e is the electron charge, 1.6×10^{-19} coulombs; and Δf is the bandwidth in cycles per second.

When the diode noise current is made to pass through a resistance R, the available noise power is the sum of (1) the available thermal noise power from the resistor P_{AT}, and (2) the additional available noise power caused by the noise from the diode P_{AD}. The available thermal noise power from any resistor P_{AT} is expressed as follows:

$$P_{AT} = KT\,\Delta f \tag{65}$$

where K is Boltzmann's constant (1.37×10^{-23} joules per degree), and T is temperature in deg K.

The available power caused by the noise from the diode is expressed as

$$P_{AD} = R\,\overline{i^2}/4 \tag{66}$$

or, using Eq. (64)

$$P_{AD} = eIR\,\Delta f/2$$

The ratio of these two available powers is the excess noise ratio of the noise-source diode plus the resistor. For the diode-resistor combination, the excess noise ratio N_S is

$$N_S = eIR/2KT \tag{67}$$

When the temperature T is the standard temperature T_0 (290 K), the numerical result is:

$$N_S = 20IR \tag{68}$$

because the quotient $e/2KT_0$ then gives the value:

$$\frac{e}{2KT_0} = \frac{1.6 \times 10^{-19}}{2 \times 1.37 \times 10^{-23} \times 290}\left(\frac{\text{coulombs}}{\text{joules}}\right)$$

$$= 20\,\frac{\text{ampere-seconds}}{\text{watt-seconds}}$$

$$= 20 \text{ reciprocal volts}$$

The value of the excess noise ratio N_S depends on the value of the resistor R connected to the diode. However, the excess noise ratio is the ratio of two available powers, so a loss-free transformer can be connected across the diode-resistor combination used as a noise source in order to transform the terminal impedance of the noise source to any desired value. The use of such a transformer would not change the value of either of the two available powers. Consequently, the excess noise ratio N_S at the output port of this transformer has the same value as the excess noise ratio of the diode-resistor combination. The assembly of the temperature-limited diode, the resistor and the transformer, with an associated power supply and meter, constitutes a noise generator.

Suppose this generator to be connected to the input port of an amplifier with a noise factor F. When the diode filament is cold, the input termination is passive and the total noise power output P_o is proportional to the noise factor F as follows:

$$P_o = CF \tag{69}$$

where C is the constant of proportionality. Alternatively, this equation may be written as follows:

$$P_o = C + C\,(F-1) \tag{70}$$

In this case the first term C represents the part of the output noise power coming from the thermal noise of the input termination, and the second term C (F-1) represents the part of the output noise power originating in the amplifier.

When the noise diode is turned on, the noise power from the input termination is increased by the excess noise ratio N_S; the total noise power P_1 can be expressed as follows:

$$P_1 = C\,(1 + N_S) + C\,(F-1)$$

$$= C\,(F + N_S) \tag{71}$$

The ratio of the two values of noise power output is

$$P_1/P_o = 1 + N_S/F \tag{72}$$

Solving for the noise factor F gives:

$$F = N_S \Big/ \left(\frac{P_1}{P_o} - 1\right) \tag{73}$$

The usual way of using a noise diode is to increase its filament temperature until the noise output of the amplifier under test is doubled. Then the denominator of Eq. (73) becomes unity and the numerator can be found from Eq. (68); the noise factor F may then be expressed as

$$F = 20IR \tag{74}$$

where F is the noise factor (power ratio), R is the resistance connected to diode in ohms, and I is the dc current in amperes required to double the noise power output.

In decibels,

$$F\,(\text{db}) = 10 \log (20IR) \tag{75}$$

The dc current meter used with a temperature-limited diode noise generator is usually provided with a scale calibrated in decibels, in accordance with Eq. (75). The scale marks in this case actually represent the excess noise ratio N_S, expressed in decibels. Eq. (73) can be used to determine the noise factor from experimental data giving noise power output ratios of any magnitude. In decibel form,

$$F\,(\text{db}) = N_S\,(\text{db}) - 10 \log \left(\frac{P_1}{P_o} - 1\right) \tag{76}$$

Some types of noise generators give only a fixed value of excess noise ratio. For example, the argon gas-discharge tube has an excess noise power ratio N_S of

34, or 15.3 db. For such generators, it is necessary to use relations of the type shown by Eq. (73) or Eq. (76) to find the noise factor. When measuring a noise factor, it is necessary that an amplifier be used which has enough gain to provide for a measurable output of noise level. Because the amplifier itself has a noise factor greater than unity, determination of the noise factor of a two-port requires that a method of separating the amplifier noise from the two-port noise be developed.

In the arrangement shown in Fig.10, the two-port (Y) with input termination Y_1 may be considered as a noise generator. When Y_1 is a passive termination at standard temperature, the available noise power P_A from the two-port (Y) is expressed as follows:

$$P_A = AGKT_0F_1\,\Delta f \tag{77}$$

where AG is the available gain of (Y), K is Boltzmann's constant, T_0 is standard temperature, F_1 is the noise factor of (Y), and Δf is the bandwidth in cycles per second.

The available noise power P_o from a passive circuit substituted for (Y) and Y_1 is $P_o = KT_0\,\Delta f$.

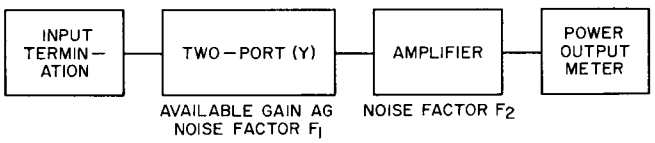


Figure 10. Noise Factor for Cascaded Amplifiers

The excess noise ratio N_S for the two-port (Y) is expressed as follows:

$$N_S = \frac{P_A}{P_o} - 1$$

$$= AGF_1 - 1$$

$$= AG + AG\,(F_1-1) - 1 \tag{78}$$

The first term AG in Eq. (78) represents the part of the excess noise ratio accounted for by the noise from the passive input termination Y_1. The noise output from the amplifier can be obtained from Eq. (78) and Eq. (71) as follows:

$$P_1 = C\left[F_2-1 + AG + AG\,(F_1-1)\right] \tag{79}$$

where C is a constant of proportionality.

Because the output noise power P_S caused by the noise from Y_1 is now given by $P_S = CAG$, the over-all noise factor F is, by definition, $F = P_1/P_S$; or $F = F_1 + (F_2 - 1)/AG$.

Solving for F_1 gives:

$$F_1 = F - \frac{F_2-1}{AG} \tag{80}$$

where F_1 is the noise factor of the two-port (Y) with the input termination Y_1, AG is the available gain of the two-port (Y) with the input termination Y_1, F_2 is the noise factor of the amplifier with the input termination equivalent to the output admittance of the two-port (Y) when the input termination of (Y) is Y_1, and F is the over-all noise factor of the system made up of the amplifier, the two-port (Y), and the input termination Y_1.

The noise factor of a two-port is a function of the input termination used. For any two-port, there is one value of input termination admittance which results in a minimum value for the noise factor. The relation between the noise factor and the input termination admittance is given by the following equation:

$$F = F_o + R_N G_o \frac{|Y-Y_o|^2}{GG_o} \tag{81}$$

where F is the noise factor of the two-port when the input termination admittance is Y, F_o is the minimum noise factor of the two-port, Y_o is the input admittance required to obtain the minimum noise factor F_o, G is the real part of Y (conductance), G_o is the real part of Y_o (conductance), and R_N is the equivalent noise resistance of the two-port.

Eq. (81) can be developed by treating the noise of a two-port as if it came from a current generator producing a current I_N and a voltage generator producing a voltage V_N, connected to the input port of an otherwise equivalent noise-free two-port, as shown in Fig. 11. Fig. 11A shown the actual circuit and Fig. 11B shows the noise-free two-port with noise generators added.

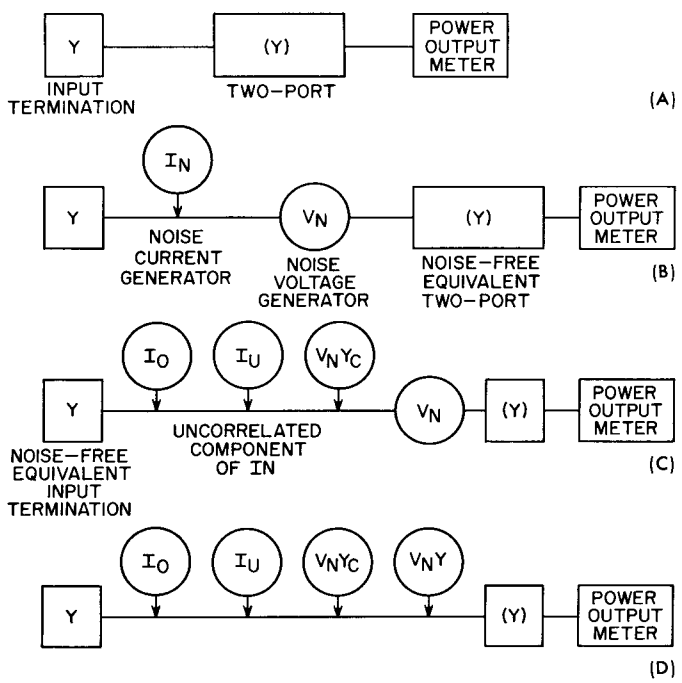


Figure 11. Equivalent Noise Generators

Let the current I_N consist of two components: (1) a component I_u which is not correlated with the voltage V_N and (2) a component I_c which is correlated with V_N. Let the component I_c be represented as the product of the voltage V_N and a complex number Y_c.

Because the quantity Y_c has the dimension of admittance, it will be referred to as the correlation admittance.

Although the voltage V_N and the current I_u are considered to have the same (average) frequency, a measurement of the phase angle between V_N and I_u at a time T_1 would not permit prediction of the phase angle between these quantities at another time T_2. A consequence of this situation is that the average value of the product of V_N and the conjugate of I_u is zero.

$$V_N I_u^* = 0 \tag{82}$$

The average value of the product of V_N and the correlated component I_c of I_N has a value:

$$\overline{V_N I_c^*} = \overline{V_N (V_N^* Y_c^*)} = \overline{V_N^2}\, Y_c^* \tag{83}$$

Consider the noise from the input termination Y as if derived from a current generator producing a current I_0. The current I_0 will not be correlated with any other noise component. Fig. 11C shows the noise generators I_0, I_u, $V_N Y_c$, and V_N.

Then consider the voltage V_N to be produced by a current generator with output $V_N Y_0$ connected in parallel with the other current generators, as shown in Fig. 11D.

The total equivalent noise input current I is expressed as follows:

$$I = I_0 + V_N (Y_0 + Y_c) + I_u \tag{84}$$

The currents $V_N Y_0$ and $V_N Y_c$ can be combined because these two currents are correlated.

The total mean-square noise input current $\overline{I^2}$ is

$$\overline{I^2} = \overline{I_0^2} + \overline{V_N^2}\,|Y_0 + Y_c|^2 + \overline{I_u^2} \tag{85}$$

Because the current generators of Fig. 11 have been chosen to represent all the noise sources of the two-port (Y) and the input termination Y, the noise power output will be proportional to the total mean-square noise input current $\overline{I^2}$. Consequently, when the input termination Y is a passive network at standard temperature T_0, the noise factor F is the ratio of the total current $\overline{I^2}$ to the component $\overline{I_0^2}$ from the input termination, and may be expressed as follows:

$$F = 1 + \frac{\overline{V_N^2}}{\overline{I_0^2}}\,|Y + Y_c|^2 + \frac{\overline{I_u^2}}{\overline{I_0^2}} \tag{86}$$

The mean-square short-circuit current $\overline{I_0^2}$ from a passive network at temperature T_0, with admittance Y = G + jB at frequency f, in a small frequency band Δf centered at frequency f is:

$$\overline{I_0^2} = 4\,K\,T_0\,G\,\Delta f \tag{87}$$

An equivalent noise resistance R_N for the two-port (Y) is defined by the equation

$$\overline{V_N^2} = 4\,K\,T_0\,R_N\,\Delta f \tag{88}$$

and an uncorrelated noise conductance G_U by the equation

$$\overline{I_u^2} = 4\,K\,T_0\,G_U\,\Delta f \tag{89}$$

Combining Eqs. (87), (88), and (89) with Eq. (86) gives

$$F = 1 + \frac{R_N}{G}\,|Y + Y_c|^2 + \frac{G_u}{G} \tag{90}$$

or

$$F = 1 + \frac{R_N\left[(G + G_c)^2 + (B + B_c)^2\right] + G_u}{G} \tag{91}$$

An optimum value Y_0 for the input termination

$$Y_0 = G_0 + jB_0 \tag{92}$$

exists, giving a minimum value F_0 for the noise factor.

It is evident that the optimum value of B_0 is

$$B_0 = -B_c \tag{93}$$

then the noise factor F_B minimized with respect to B is

$$F_B = 1 + 2\,R_N G_c + R_N G + \frac{R_N G_c^2 + G_u}{G} \tag{94}$$

The minimum value F_0 is obtained for the value G_0 for G which equalizes the last two terms of Eq. (94).

Consequently,

$$R_N G_0^2 = R_N G_c^2 + G_u \tag{95}$$

$$G_0 = \sqrt{G_c^2 + G_u/R_N} \tag{96}$$

and

$$F_0 = 1 + 2\,R_N G_c + 2\,R_N G_0 \tag{97}$$

Subtracting Eq. (97) from Eq. (91) gives:

$$F - F_0 = \frac{R_N G^2 + R_N G_c^2 + G_u}{G} - 2\,R_N G_0 + R_N \frac{(B + B_c)^2}{G} \tag{98}$$

Substitution of values from Eq. (93) and Eq. (95) in Eq. (98) gives:

$$F - F_0 = \frac{R_N G^2 + R_N G_0^2 - 2\,R_N G_0 G}{G} + R_N \frac{(B - B_0)^2}{G} \tag{99}$$

$$F = F_0 + \frac{R_N\left[(G - G_0)^2 + (B - B_0)^2\right]}{G} \tag{100}$$

$$= F_o + \frac{R_N \left| Y - Y_o \right|^2}{G} \qquad (101)$$

An alternative form for Eq. (101) is

$$F = F_o + R_N G_o \frac{\left| Y - Y_o \right|^2}{G G_o} \qquad (102)$$

This form is identical with Eq. (81)

The terms and groups F, F_o, $R_N G_o$ and $Y - |Y_o|^2 / GG_o$ in Eq. (102) are all invariant with insertion of loss-free passive two-ports between the two-port (Y) and the input termination Y.

Measurement of minimum noise factor, optimum input termination admittance, and equivalent noise conductance are made most conveniently with direct-reading noise-factor measuring equipment such as the Hewlett-Packard model 340 noise-figure meter. Equipment may be set up as shown in Fig. 12. The transformer and the admittance measuring equipment are subject to the same considerations discussed previously. The following procedure is used with this set-up:

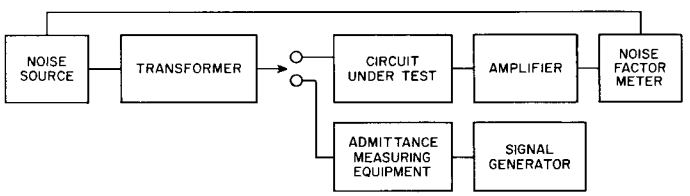


Figure 12. Measurement of Noise Factor

1. Adjust the transformer for a minimum noise factor reading. Record this reading.

2. Switch to the admittance-measuring circuit and record the conductance and susceptance.

3. Adjust the transformer to obtain a chosen higher noise factor reading.

4. Repeat Step. 2.

Values for F_o, G_o, and B_o for use in Eq. (81) are obtained from Steps 1 and 2 and values for F, G, and B are obtained in Steps 3 and 4. Insertion of these values in Eq. (81) permits solution for the equivalent noise resistance R_N and the product $R_N G_o$.

Some of the results obtained previously show that there are restrictions applying to acceptable values obtained by the above procedures. Thus, Eq. (97) is

$$F_o = 1 + 2 R_N G_c + 2 R_N G_o \qquad (97)$$

therefore,

$$\frac{F_o - 1}{2 R_N G_o} - 1 = \frac{G_c}{G_o} \qquad (103)$$

Eq. (96) is

$$G_o = \sqrt{G_c^2 + G_u / R_N} \qquad (96)$$

therefore,

$$\frac{G_o}{G_c} = \sqrt{1 + \frac{R_N G_u}{(R_N G_c)^2}} \qquad (104)$$

The uncorrelated noise conductance G_u, as defined by Eq. (89), is necessarily positive, so the ratio G_o/G_c must be equal to or greater than unity and the ratio G_c/G_o must be equal to or less than unity. However, the correlation conductance G_c may be positive or negative. Consequently, the ratio $(F_o-1)/2R_N G_o$ must have values in the range 0 to +2; if values greater than 2 are obtained, something is wrong with the calibration of the equipment or with the recorded data. A value of unity signifies no correlation between the equivalent noise resistance (voltage) and the equivalent noise conductance (current) when the noise is referred to the input port as in the preceding discussion.

Corrections for Lead Inductances and Capacitances

The internal electrode structure of a tube may be considered as a two-port even though the electrodes themselves are not accessible. The problem is to find the signal-frequency voltages between the tube electrodes and the currents into the tube electrodes from the results of measurements made at the accessible terminals of the assembly of tube, socket, and filters. For an example of procedures which might be used, consider the circuits of Fig. 13.

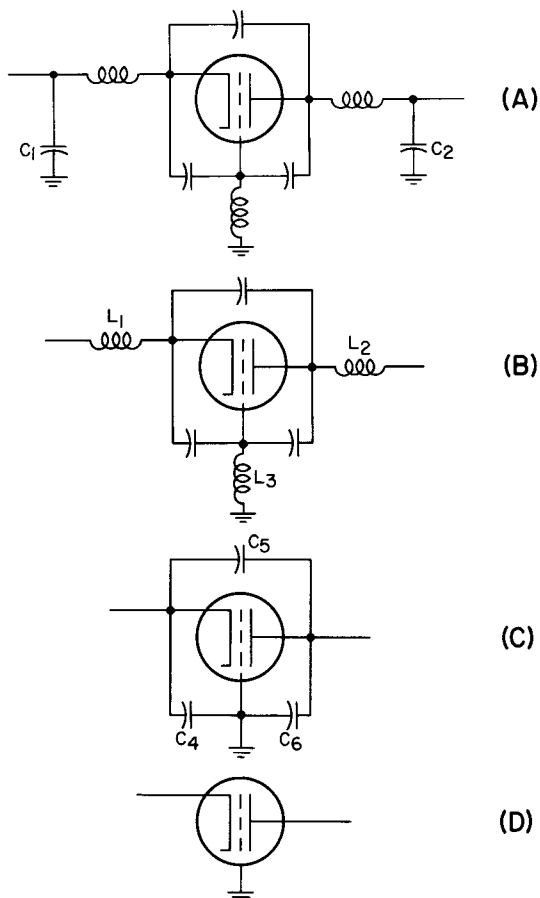


Figure 13. Correction for Lead Inductances and Capacitances

First, assume that the short-circuit admittances have been found for the circuit (A) of Fig. 13. The capacitances C_1 and C_2 might represent capacitances between socket terminals and ground. At an angular frequency ω radians per second, C_1 contributes a susceptance component $J\,\omega C_1$ to the short-circuit input susceptance Y_{11}, and C_2 contributes a susceptance component $J\omega C_2$ to the short-circuit output susceptance Y_{22}. Subtracting these components from Y_{11} and Y_{22} gives the values applicable to the circuit (B) of Fig. 13

The short-circuit admittance equations for this circuit may be written as follows:

$$I_{1b} = V_{1b}\,Y_{11b} - V_{2b}\,Y_{12b}$$

$$I_{2b} = -V_{1b}\,Y_{21b} - V_{2b}\,Y_{22b}$$

Solving these equations for the voltages gives the open-circuit impedance equations for the same circuit.

Based on the equation

$$D_b = Y_{11b}\,Y_{22b} - Y_{12}\,Y_{21b}$$

and subject to the condition $D \neq 0$, the open-circuit impedance equations are written as follows:

$$V_{1b} = I_{1b}\,Z_{11b} + I_{2b}\,Z_{12b}$$

$$V_{2b} = I_{1b}\,Z_{21b} + I_{2b}\,Z_{22b}$$

where

$Z_{11b} = Y_{22b}/D_b$ is the open-circuit input impedance

$Z_{12b} = Y_{12b}/D_b$ is the open-circuit feedback impedance

$Z_{21b} = Y_{21b}/D_b$ is the open-circuit forward impedance

$Z_{22b} = Y_{11b}/D_b$ is the open-circuit output impedance

The inductances L_1 and L_2, shown in Fig. 13B, contribute reactance components $J\,\omega L_1$ and $J\,\omega L_2$ to the open-circuit input and output reactances, respectively, and the inductance L_3 contributes a component $J\,\omega L_3$ to each of the four open-circuit reactances. Subtracting out these components gives the values applicable to Fig. 13C.

$$Z_{11c} = Z_{11b} - J\omega(L_1 + L_3)$$

$$Z_{12c} = Z_{12b} - J\,\omega L_3$$

$$Z_{21c} = Z_{21b} - J\,\omega L_3$$

$$Z_{22c} = Z_{22b} - J\omega(L_2 + L_3)$$

The open-circuit voltage equations for the circuit of Fig. 13C may then be solved for the currents. The solution gives the short-circuit admittance coefficients for the circuit of Fig. 13C. The capacitances C_4 and C_6 contribute susceptances $J\,\omega\,C_4$ and $J\,\omega\,C_6$ to the short-circuit input and output admittances, and the capacitance C_5 contributes a susceptance $J\,\omega\,C_5$ to each of the four admittances. Subtracting these components gives the short-circuit admittance coefficients of Fig. 13D, which represents the internal electrode structure of the tube.

The applications of the indicated procedures depends on knowing or estimating the values of the inductances and capacitances involved. It is not difficult to measure the capacitances between the terminals of a socket or the capacitances between each terminal and the plate to which the socket is mounted. Capacitances between the leads of a tube can be found by carefully removing the envelope and the electrode structure of a tube and then measuring the desired capacitances, or by obtaining the desired tube parts and measuring the capacitances. In many cases an inductance value can be determined with sufficient accuracy by making a series of measurements of susceptances at a number of frequencies with a cold tube in a relatively simple circuit and determining the amount of inductance necessary to account for the observed variation of susceptance with frequency. Because any measurement of inductance actually must include all of the self-inductances and mutual inductances of the components of a closed circuit, it is still necessary to make an estimate of the values of inductance to be assigned to the various leads in a test circuit after the total value for the inductance of a loop containing these leads has been determined.

Estimated values of inductance determined from the lengths of leads in a tube or a socket may be used when data from measurements are not available. The inductance of a lead which is part of a circuit is determined principally by the length of that lead and is affected to a smaller degree by the lead diameter and by distances of that lead from the leads completing the circuit.

Fig. 14 shows a rectangular loop formed from a round wire. When the width of the loop x is small compared to the length y/2 and the wire radius r is small compared to the width x, the inductance of the loop is

$$L = 2y \log_e \frac{x}{r}$$

where L is the inductance, in millimicrohenries, and y is twice the length of the loop, in centimeters.

When the ratio x/r is equal to 20, the natural logarithm of this ratio is 3; therefore, the inductance per centimeter of length of wire is approximately 6 millimicrohenries. For a logarithm value of 2.5, the ratio x/r is approximately 12; this ratio gives an inductance value of 5 millimicrohenries per centimeter of length. Because ratios of spacing-to-radius between 12 and 20, corresponding to ratios of spacing-to-diameter between 6 and 10, are typical values for the leads in vacuum tubes and sockets, good approximations to actual values are obtained by assuming that the inductances of the tube and socket leads amount to 5 or 6 millimicrohenries per centimeter of length. When two or more leads are connected to a single tube electrode, it is generally permissible to divide the estimated inductance per lead by the number of leads.

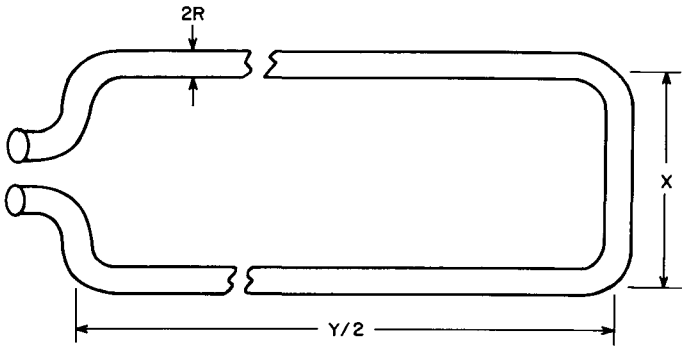


Figure 14. Rectangular Loop formed from Round Wire

The total length of a lead from a tube element to the chassis in the case of a miniature tube of standard construction is about two centimeters; the inductance per lead is about 10 or 12 millimicrohenries. This estimate is made for a tube in a socket of standard construction and a very short connection between socket terminal and chassis.

LARGE-SIGNAL CONSIDERATIONS

The large-signal performance of a small-signal amplifier must be discussed on the basis that the tube is a nonlinear transducer. When signals reaching a tube are small, they are amplified without distortion of the modulation envelopes and without appreciable production of harmonies. Moreover, when several small signals of different frequencies reach the input of a linear amplifier, they are amplified without interaction and, therefore, may be subsequently separated by frequency-selective circuits.

Consider an amplifier tube having a bias voltage applied so it operates near or beyond cutoff. Because the forward transfer admittance is low or zero under these conditions, there is little or no amplification of small signals. However, the peaks of a large signal cause considerable current to flow; consequently, the peaks are amplified. For example, a television signal would be distorted because the synchronizing pulses and the signals corresponding to the black picture portions pass through the tube, but the signals corresponding to lighter portions are lost. Moreover, the picture signal passed during the lighter parts of the picture may be so low that the carrier needed for intercarrier sound is interrupted, which would produce a bad buzz in the sound.

When a tube is operated with very low bias voltage and a large signal is applied to the tube input, the effect is to reduce the plate-circuit peaks because of the limiting action of grid-current flow, although amplification of smaller signals takes place normally. For a television signal, this action would result in possible loss of synchronizing signals and compression of the dark levels.

A tube biased near cutoff and subjected to a strong undesired signal can produce cross modulation. In an extreme case, the undesired signal can, in effect, turn the tube on and off to permit passage of the desired signal when the amplitude of the undesired signal is large and to suppress the desired signal when the amplitude of the undesired signal is small.

In practical situations, trouble is experienced with signals which are not large enough to carry the grid voltage of a tube beyond cutoff or into grid current. Estimates of the signal-handling capacities of tubes can be made by measuring the degree of nonlinearity of their transfer characteristics and then expressing the results in terms of the signal amplitude required to produce a given amount of interference.

The characteristic in Fig. 15 shows plate current as a function of grid voltage; for any grid voltage, for example V_c, which gives a non-zero plate current, a power series can be used to represent the curve. The coefficients in this series are proportional to the successive derivatives of the plate current with respect to the grid voltage. Let V_s be a signal voltage added to the grid voltage V_c; then, defining the signal current I_s as the difference between the current for $V = V_c + V_s$ and the current for $V = V_c$, the power-series equation is

$$I_s = V_s \frac{dI_b}{dE_c} + \frac{V_s^2}{2} \frac{d^2I_b}{dE_c^2} + \frac{V_s^3}{2+3} \frac{d^3I_b}{dE^2} + \cdots \frac{V_s^m}{m!} \frac{d^mI_b}{dE^m} \cdots$$

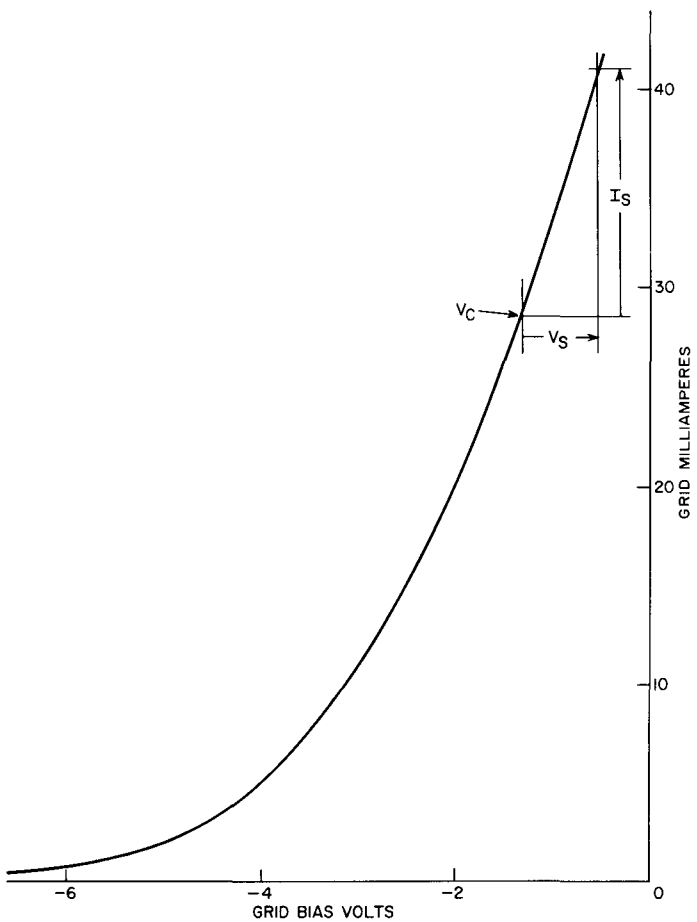


Figure 15. Plate Current as a function of Grid Bias

But the first derivative of plate current with respect to grid voltage is the transconductance G_m of the tube. The second derivative of the plate current, consequently, is the first derivative of the transconductance, etc. Denoting the successive derivatives of the transconductance by G_m', G_m'', and the (N-1)st derivative by $G_m^{<N-1>}$, and using a summation notation for compactness, the equation for the signal current can be written as:

$$I_s = \sum_{N=1}^{\infty} \frac{V_s{}^N}{N!} G_m^{<N-1>} \tag{105}$$

For this equation to be valid, the signal voltage V_s must not be so large as to take the grid voltage past cutoff or past any region where the plate current and its derivatives cannot be considered as continuous. This condition excludes signal excursions into regions of positive grid voltage when large grid resistors are used with a tube.

Let the signal voltage V_s be the sum of two sinusoidal voltages with peak magnitudes V_1 and V_2 and angular frequencies ω_1 and ω_2. Cosine functions may be used to represent these voltages, as follows:

$$V_s = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t$$

Substitution of equivalent exponential functions for the cosine functions gives:

$$V_s = \frac{V_1}{2} e^{j\omega_1 t} + \frac{V_1}{2} e^{-j\omega_1 t} + \frac{V_2}{2} e^{j\omega_2 t} + \frac{V_2}{2} e^{-j\omega_2 t} \tag{106}$$

The equation representing the sum of a number of terms raised to a power, in summation notation, is:

$$(p + q + \ldots w)^N = \sum_{\substack{a, b, c \ldots j \\ a + b + \ldots j = N}} \frac{N!}{a!b! \ldots j!} p^a q^b \ldots w^j$$

Consequently, the equation for the Nth power of the signal voltage can be written as:

$$V_s{}^N = \sum_{\substack{a, b, c, d. \\ a + b + c + d = N}} \frac{N!}{a!b!c!d!} \left(\frac{V_1}{2}\right)^{a+b} \left(\frac{V_2}{2}\right)^{c+d} e^{j[(a-b)\omega_1 + (c-d)\omega_2]t} \tag{107}$$

After substituting Eq. (107) in Eq. (105), the equation for signal current becomes:

$$I_s = \sum_N \sum_{\substack{a, b, c, d \\ a + b + c + d = N}} \frac{G_m^{<N-1>}}{a!b!c!d!} \left(\frac{V_1}{2}\right)^{a+b} \left(\frac{V_2}{2}\right)^{c+d} e^{j[(a-b)\omega_1 + (c-d)\omega_2]t} \tag{108}$$

Consider the case for which the plate circuit of the tube is tuned to the angular frequency ω_1. The only components of the signal current giving outputs at this frequency are those for which the exponent of e is $j\omega_1$ or $-j\omega_1$. The conditions giving these values are:

$$a - b = \pm 1$$

$$c - d = 0$$

For these conditions, the sum of a and b must be odd and the sum of c and d must be even, so the sum of all four terms must be odd. This means that the index number N must be odd and, consequently, terms with the angular frequency ω_1 are derived from the terms containing the transconductance G_m and its even derivatives G_m'', G_m^{iv}, etc.

For N = 1, the only permissible values for a, b, c, and d are

a	b	c	d
1	0	0	0
0	1	0	0

For N = 3, the permissible values are

a	b	c	d
1	0	1	1
0	1	1	1
2	1	0	0
1	2	0	0

For N = 5, the permissible values are

a	b	c	d
1	0	2	2
0	1	2	2
2	1	1	1
1	2	1	1
3	2	0	0
2	3	0	0

If I_1, the plate-current component, has an angular frequency ω_1, the equation containing the contributions from terms including N = 5 is obtained by substituting the permissible values for a, b, c, and d in Eq. (108); the resulting equation for I_1 is

$$I_1 = (e_1 e^{j\omega_1 t} + e_1 e^{-j\omega_1 t}) \left[G_m \frac{V_1}{2} + G_m'' \left\{ \frac{V_1}{2} \left(\frac{V_2}{2}\right)^2 + \frac{1}{2}\left(\frac{V_1}{2}\right)^3 \right\} + G_m^{iv} \left\{ \frac{1}{4} \frac{V_1}{2} \left(\frac{V_2}{2}\right)^4 + \frac{1}{2}\left(\frac{V_1}{2}\right)^3 \left(\frac{V_2}{2}\right)^2 + \frac{1}{12}\left(\frac{V_1}{2}\right)^5 \right\} + \ldots \right]$$

This relationship is equivalent to:

$$I_1 = \cos \omega_1 t \left[G_m V_1 + G_m'' \left(V_1 \frac{V_2{}^2}{4} + \frac{V_1{}^3}{8} \right) + G_m^{iv} \left(\frac{V_1 V_2{}^4}{64} + \frac{V_1{}^3 V_2{}^2}{32} + \frac{V_1{}^5}{192} \right) + \ldots \right]$$

or

$$I_1 = G_m V_1 \cos \omega_1 t \left[1 + \frac{G_m''}{G_m} \left(\frac{V_2^2}{4} + \frac{V_1^2}{8} \right) + \frac{G_m^{iv}}{G_m} \left(\frac{V_2^4}{64} + \frac{V_1^2 V_2^2}{32} + \frac{V_1^4}{192} \right) + \cdots \right] \quad (109)$$

The voltage output is proportional to the plate-current component I_1. The voltage gain G is defined as a quantity proportional to the quotient of the current component I_1 divided by the voltage ($V_1 \cos \omega t$). At low values of V_1 and V_2 the small-signal gain G_0 is proportional to G_m; therefore, the gain G is

$$G = G_0 \left[1 + \frac{G_m''}{G_m} \left(\frac{V_2^2}{4} + \frac{V_1^2}{8} \right) + \frac{G_m^{iv}}{G_m} \left(\frac{V_2^4}{64} + \frac{V_1^2 V_2^2}{32} + \frac{V_1^4}{192} \right) + \cdots \right] \quad (110)$$

The following statements apply to the voltage gain:

1. For small values of V_1 and V_2, the gain is proportional to the product of the signal-frequency voltage V_1 and the transconductance G_m.

2. When the undesired signal voltage V_2 is zero or small, but the desired signal voltage V_1 is large, the gain is modified by terms proportional to the products of (a) even powers of the voltage V_1 and (b) the quotients of even-order derivatives of the transconductance divided by the transconductance.

3. When the undesired signal voltage V_2 is large, the gain is modified by terms proportional to the products of (a) even powers of the voltage V_2, (b) possibly, even powers of the voltage V_1, and (c) the quotients of even-order derivatives of the transconductance divided by the transconductance.

4. When either of the signals is amplitude modulated (i.e., V_1 or V_2 varies periodically with time) and this signal is large, the gain varies with time.

   (a) In the case of modulation of the desired signal, the result is a nonlinear relation between the modulation of the output signal and the input signal with an increase or decrease in the relative depth of modulation.
   (b) In the case of modulation of the undesired signal, the result is a modulation of the output signal, which is a function of the modulation of the undesired signal, i.e., cross modulation.

The characteristic in Fig. 16 shows transconductance as a function of grid voltage. This curve can be obtained from the characteristic showing plate current as a function of grid voltage by differentiation, or it can be obtained by plotting data from direct measurement of transconductance.

Positive values of the second derivative of transconductance with respect to grid voltage (or third derivative of the plate current) result in a concave-upward curvature of this characteristic; negative values result in a concave-downward curvature. At the inflection point, the second derivative is zero.

Two cases illustrating the effect of the application of a large signal voltage are shown in Fig. 16. For the more negative grid voltage, with the indicated value of signal applied, the transconductance averaged with respect to time is greater than the transconductance for the grid bias corresponding to the selected operating point. The second derivative of transconductance is positive at this point. For the less negative grid voltage, the second derivative is negative and the average transconductance is less than the operating-point transconductance. In either case, the development of cross modulation can be illustrated by considering the effects of changing the maximum value of the signal voltage V_s. When this voltage is very small, the gain obtained for a small desired signal (say, V_1) is proportional to the transconductance G_m at the operating point. For larger values of V_s, the gain for V_1 is proportional in each instance to the average value of G_m for the value of V_s; therefore, the modulation envelope for the output signal derived from V_1 corresponds to the curve of average transconductance as a function of time derived from the modulation envelope of the signal V_s.

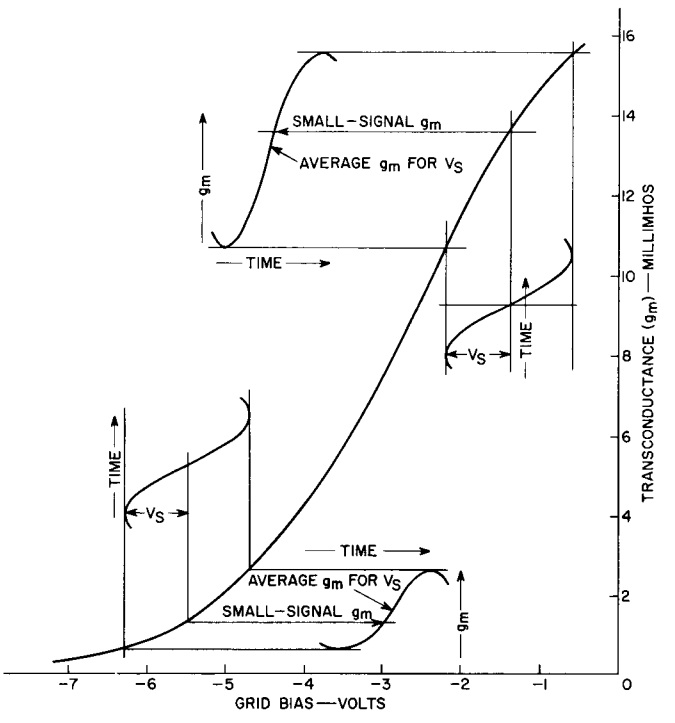


Figure 16. Transconductance as a function of Grid Bias

The second derivative at a point on a curve such as that of Fig. 16 may be determined by subtracting the G_m at the selected point from the mean of the two G_m values for voltages displaced by equal amounts (e.g., by V_s) from the selected voltage. Then divide this result by one-half the square of the displacement V_s. The

second derivative is the limiting value of this result as the displacement V_s is made small. In equation form,

$$G_m'' = \lim_{V_s \to 0} \frac{\frac{G_m(V_c + V_s) + G_m(V_c - V_s)}{2} - G_m(V_c)}{\frac{1}{2}V_s^2}$$

The procedure for obtaining a cross-modulation envelope by determining average transconductance values for signals of prescribed magnitude is closely related to this method of obtaining the second derivative. The procedure using average transconductance may be considered as a method of taking into account contributions from the fourth and higher derivatives of the transconductance characteristics and as a method for obtaining a result when the continuity requirements of the power-series equation are not satisfied.

The first two terms of Eq. (109), i.e.,

$$I_1 = G_m V_1 \cos \omega_1 t \left[1 + \frac{G_m''}{G_m}\left(\frac{V_2^2}{4} + \frac{V_1^2}{8}\right)\right] \quad (111)$$

give enough information about a tube characteristic to permit discussion of various characteristic curve shapes as they affect cross modulation and modulation distortion.

The use of this equation for determining cross modulation may be illustrated by substituting an expression representing an amplitude-modulated signal for the voltage V_2; thus,

$$V_2 = V(1 + M \cos pt)$$

where M is the modulation coefficient of the undesired signal.

Squaring this expression gives:

$$V_2^2 = V^2\left(1 + \frac{M^2}{2} + 2M \cos pt + \frac{M^2}{2}\cos 2pt\right)$$

Substituting this result in Eq. (111) gives:

$$I_1 = G_m V_1 \cos \omega_1 t \left[1 + \frac{G_m''}{G_m}\left(\frac{V^2}{4} + \frac{V^2}{8}M^2 + \frac{V_1^2}{8}\right) + \frac{G_m''}{G_m}\frac{V^2}{2}M \cos pt + \cdots\right] \quad (112)$$

When M' is the first approximation to the modulation coefficient of the signal-frequency current I_1, obtained from Eq. (112), then

$$M' = M\frac{G_m''}{G_m}\frac{V^2}{2} \quad (113)$$

Solving for the voltage V gives:

$$V = \left(\frac{G_m}{G_m''}\right)^{1/2}\left(2\frac{M'}{M}\right)^{1/2} \quad (114)$$

When the degree of interference which can be accepted is specified, the magnitude of the tolerable interfering voltage can be found from Eq. (114). For example, if the allowable cross-modulation ratio (M'/M) is specified as 2 per cent, the tolerable voltage is

$$V_{(2\text{ per cent})} = \sqrt{0.04}\left(\frac{G_m}{G_m''}\right)^{1/2} = 0.2\left(\frac{G_m}{G_m''}\right)^{1/2}$$

For 10% cross modulation,

$$V_{(10\text{ per cent})} = 0.45\left(\frac{G_m}{G_m''}\right)^{1/2}$$

When a curve of transconductance as a function of grid voltage is available for a tube, the second derivative (G_m'') of the transconductance can be found and consequently the quantity $(G_m/G_m'')^{1/2}$ used in Eq. (114) can be derived.

It is generally necessary to vary the gain of the first tube in a receiver by adjustment of the control-grid bias, to accommodate the receiver to the various levels of incoming signal that may be encountered. The voltage gain from the rf amplifier stage is proportional to the transconductance G_m of the tube used in this stage. Fig. 17 is a curve of the quantity $(G_m/G_m'')^{1/2}$ plotted against the relative gain in decibels for a tube having characteristics approximately the same as those for curves of Figs. 15 and 16. This curve shows that a tube with the characteristics illustrated can accept a fairly large interfering signal voltage when the gain is high, but the tolerance to interfering signals is greatly reduced for the low-gain conditions. For 2 per cent cross modulation, the tolerable interfering signal for a gain 22 db below the gain at zero bias would be 0.2 volts.

Eq. (109) can also be used to determine the distortion of the modulation of the desired signal. For this purpose, let the undesired signal voltage V_2 be zero and substitute an expression representing an amplitude-modulated signal for the voltage V_1, thus:

$$V_1 = V(1 + M \cos pt)$$

$$V_1^2 = V^2\left(1 + \frac{M^2}{2} + 2M \cos pt + \frac{M^2}{2}\cos 2pt\right)$$

then, from Eq. (111)

$$I_1 = G_m V \cos \omega t\,(1 + M \cos pt)\left[1 + \frac{G_m''}{G_m}V^2\left(\frac{1}{8} + \frac{M^2}{16}\right) + \frac{G_m''}{G_m}V^2\left(\frac{M}{4}\cos pt + \frac{M^2}{16}\cos 2pt\right)\right] \quad (115)$$

or

$$I_1 = G_m V \cos \omega t\left[1 + \frac{G_m''}{G_m}V^2\left(\frac{1}{8} + \frac{3}{16}M^2\right)\right] + G_m V \cos \omega t \cos pt\left[M\left(1 + \frac{3}{8}\frac{G_m''}{G_m}V^2 - \frac{1}{32}\frac{G_m''}{G_m}V^2M^2\right)\right]$$

$$+ G_m V \cos \omega t \cos 2 pt \left[\frac{3}{16} \frac{G_m''}{G_m} V^2 M^2\right]$$

$$+ \cdots$$

or, approximately

$$I_1 = G_m V \left(1 + \frac{G_m''}{G_m} \frac{V^2}{8}\right) \cos \omega t \left[1 + M + \frac{1}{4} \frac{G_m''}{G_m} V^2 M \cos pt + \frac{3}{16} \frac{G_m''}{G_m} V^2 M^2 \cos 2 pt + \cdots\right]$$

The fundamental-frequency modulation coefficient M' is

$$M' = M\left(1 + \frac{1}{4} \frac{G_m''}{G_m} V^2\right) \qquad (116)$$

The second-harmonic modulation coefficient M_2' is

$$M_2' = M^2 \left(\frac{3}{16} \frac{G_m''}{G_m} V^2\right) \qquad (117)$$

The tolerable signal for a specified change in modulation or for a specified harmonic distortion is proportional to the square root of the quotient of the transconductance divided by the second derivative of the transconductance, as in the case of cross modulation. Solving Eq. (116) for the voltage V gives:

$$V = \left[4 \frac{G_m}{G_m''} \frac{M' - M}{M}\right]^{1/2}$$

If a change in modulation of 25 per cent is permitted,

$$V_{(25 \text{ per cent})} = (G_m/G_m'')^{1/2}$$

If the change in modulation is limited to 5 per cent,

$$V_{(5 \text{ per cent})} = 0.45 (G_m/G_m'')^{1/2}$$

Curves like that shown in Fig. 17 can be used to determine the allowable signal magnitudes for any chosen specification of cross modulation from an interfering signal or distortion of modulation of the desired signal.

The signals likely to be encountered in a television receiver can be estimated by means of very simple considerations. Suppose a transmitter emits power equally in all directions. The area of a spherical surface at a radius of 1000 meters is $4 \pi \times 10^6$ square meters. For a radiated power of 50 kilowatts, the power density of the signal at 1000 meters from the source would be $(50 \times 10^3)/(12.5 \times 10^6) = 4 \times 10^{-3}$ watts per square meter. A receiving antenna with an effective area of one square meter would thus receive this amount of power. If this power was then delivered to the grid of a tube with an input impedance of 1000 ohms, the rms voltage at the grid of this tube would be 2 volts.

Calculations taking into account reflections from the earth lead to lower estimated signal levels, but on the other hand the directivity of the transmitting antenna and the larger effective areas of many receiving antennas operate to increase the signal levels. In a metropolitan area there can be thousands of receivers at distances of less than 1000 meters from a transmitter, so it is necessary to consider signal magnitudes of the order of volts at the grid of the first tube in a receiver. On the other hand, the smallest usable signal magnitude is determined by the noise level of the receiver and may be of the order of 50 microvolts.

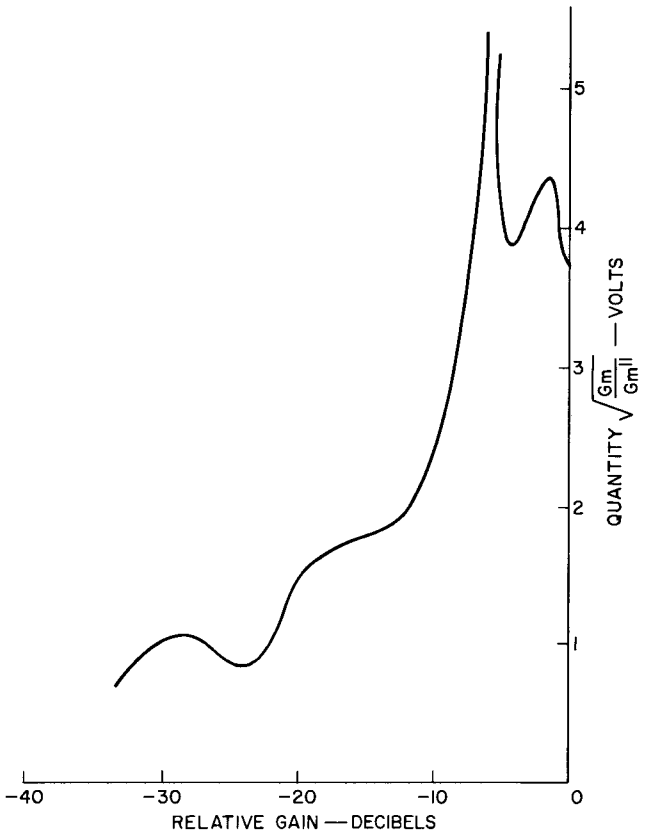


Figure 17. Characteristics for Evaluating Cross Modulation and Distortion

A television receiver for the vhf band uses a tuner consisting of an rf amplifier stage with gain adjusted by means of controlled grid bias voltage and a mixer stage with no gain control. At minimum negative bias the voltage gain of this combination may be from 30 to 50 times. Because the first if amplifier tube cannot operate properly with a signal level greater than approximately one volt, it is necessary to reduce the gain of the tuner to unity or less than unity to handle the largest probable signals. The required reduction in gain is of the order of 30 decibels.

The magnitudes of interfering signals are reduced by the selectivity provided in the tuner between the antenna and the grid of the first tube. Thus a signal which could develop a two-volt magnitude at the grid of the first tube when tuned to resonance might produce a few tenths of a volt or half a volt at the grid when the tuner is switched to another channel. The cross modulation produced by an interfering signal of a few tenths of a volt may, however, be more objectionable than the dis-

tortion of modulation of a desired signal of a much larger magnitude. The same properties of the tube characteristics are involved in either case, so attention to the cross-modulation aspect of the large-signal problem is justified.

Consider the tube represented by the characteristic curves shown in Figs. 15, 16, and 17. An immediate objection to use of this tube is the high plate current indicated for zero grid bias (Fig. 15). This current can be reduced by supplying the plate voltage for the tube through a series resistor; then the parts of the curves for the lower currents will not be changed a great deal, but the maximum transconductance will be reduced along with the maximum plate current. A new determination of the ratio of the transconductance to its second derivative (Fig. 17) would be necessary for the changed operating conditions. However, it is clear that the relative-gain scale would be shifted in accordance with the amount of reduction of the zero-bias transconductance and that the left-hand part of the curve, corresponding to low plate currents would not be affected very much. Suppose the current is limited to 25 milliamperes and the maximum transconductance is reduced from about 16 millimhos to 12 millimhos. The zero-bias gain is thus reduced by 2.5 db. The first minimum for the quantity $(G_m/G_m'')^{1/2}$ as shown in Fig. 17 would be found at about 20 db gain reduction. The allowable interfering signals for relative gain values from -20 to -30 db would be of the order of 0.2 volt at the grid of the tube.

Improvements to be sought might be higher transconductance at the maximum current value and a tolerance of larger signals. A more extended gain-control characteristic would also be desirable. The characteristics obtainable are limited, however, by the fundamental relations between the plate current, the transconductance, and the higher derivative of tube characteristics.

For consideration of cross-modulation interference, it is reasonable to assume that the probability of a given magnitude of interfering signal is independent of the magnitude of the desired signal. The desirable tube characteristic would be one for which the ratio of the transconductance to the second derivative would be the same for all values of transconductance. This condition is satisfied by a tube characteristic following an exponential law, thus:

$$I_b = I_o e^{KVg}$$

Successive derivatives are

$$G_m = KI_o e^{KVg}$$

$$G_m' = K^2 I_o e^{KVg}$$

$$G_m'' = K^3 I_o e^{KVg}$$

and consequently

$$(G_m/G_m'')^{\frac{1}{2}} = 1/K$$

but also

$$G_m/I_b = K$$

Therefore, if a value 1/K of one volt over the whole operating range is specified, the quotient G_m/I_b is thereby limited to unity; any higher quotient of G_m/I_b implies a lower tolerance to cross modulation in at least part of the operating range.

The tube design procedure for controlling the shape of the characteristic of the transconductance as a function of grid voltage is the use of nonuniform spacings of the control-grid wires. The characteristic curve obtained can be considered equivalent to that of a number of tubes with different cutoff voltages operated in parallel. Tubes with characteristics approximating exponential curves have been made in this manner. It is more common practice, however, to compromise the design to obtain a high ratio of transconductance to plate current at the zero-bias end of the characteristic. Then the interfering signal level for a specified degree of cross modulation goes through a pronounced minimum in some part of the operating range.

If a curve such as shown in Fig. 17 is given as a specification for a tube characteristic, the curves of transconductance as a function of grid voltage and plate current as a function of grid voltage can be constructed, given this curve and the plate current and transconductance for one value of grid voltage. The curve actually represents a third-order differential equation for which a unique solution is obtainable when the given values of plate current and transconductance are used as integration constants.

TUBE DESIGN CONSIDERATIONS

The design requirements for tubes giving high gain, low-noise factor, and adequate bandwidth at high frequencies are met first by reducing the tube size and the internal spacings to a minimum, and then by careful attention to a great many factors which can help or hurt the performance of a tube. Among such factors are the texture of the cathode coating, cathode current density, the uniformity of grid winding, and the lengths and arrangement of the leads.

Small size is indicated first by purely static considerations. It is generally desirable to have the resonant frequencies of the circuits represented by the lead inductances and interelectrode capacitances of a tube high compared to the desired operating frequencies; achieving this relationship means that the dimensions involved should be quite small in comparison with the free-space wavelengths corresponding to the operating frequencies. A very broad principle of scaling can be mentioned here; if a given structure meets certain standards of performance at a particular frequency, a similar structure in which all linear dimensions are reduced by a factor of two should meet the same standards at twice the frequency. Unfortunately, because scaling in this manner quickly leads to difficulties involving cathode current densities, mechanical strength of parts, and heat conduction, compromises are necessary if practical designs are to be obtained.

The equivalent diode represented by the cathode and

the first grid is the part of a tube requiring attention first. The transconductance of a triode, tetrode, or pentode is proportional to the conductance of this equivalent diode for the tube. For a parallel-plane structure, it is evident that all capacitances will be proportional to the cathode area used; it, therefore, is advantageous to make this area small. For a given cathode area and anode voltage, the conductance of a diode is inversely proportional to the square of the spacing between cathode and anode. Consequently, the numerical ratio of conductance to capacitance is inversely proportional to spacing under these conditions. In the case of a triode, the parallel-plane relations are not realized unless the grid-wire diameter and the grid-pitch distance are small in comparison to the grid-cathode spacing. This statement does not mean that closer spacings are not to be used, but it does imply that theory based on parallel-plane geometry will not be adequate to predict the characteristics of tubes with relatively small spacings.

The effective voltage in the grid plane is the principal factor controlling the ratio of transconductance to current. If the ratio of transconductance to current and the cathode area are specified, one approach to optimum design is to choose the smallest grid-wire size considered practical and then determine the pitch and spacing which will give maximum current and transconductance with this size of wire. The mechanical problems incident to obtaining and controlling very close spacings are ignored in this discussion. As for the grid-to-plate spacing, large values of it lead to low capacitance and high amplification factor, but also to requirements for high anode voltage and high power dissipation. The value chosen will depend on the relative importance of these factors in different designs. The same remarks apply to the grid-to-screen grid spacing in a tetrode or pentode.

With grounded-cathode operation, part of the short-circuit input conductance of a triode, tetrode, or pentode is determined by phase delay in the ac component of the anode current; this component is a function of the electron transit times between cathode and grid and between grid and anode. For a given effective dc voltage at the grid, the transit time for an electron moving from cathode to grid is proportional to the cathode-grid spacing; this relationship is an additional reason for keeping that spacing as small as possible. There is a part of the input conductance which is proportional to the product of grid-cathode capacitance and cathode lead inductance. Therefore, it follows that a design which keeps the grid-cathode capacitance to a minimum for a given value of transconductance tends to minimize the second part of the input conductance.

The minimum noise factor for a tube was found to be a function of the product of the equivalent noise resistance R_n and the optimum input termination conductance G_o. The theory for triodes with parallel-plane geometry indicates that R_n should be inversely proportional to the transconductance G_m, and that the optimum conductance G_o is a function of R_n and the component of noise current at the input port not correlated with that at the output port. Theory for this particular component of noise current has not been fully developed; in any event, it is known that substantial deviations from minimum values occur in actual tubes because of deviations from parallel-plane conditions and losses in dielectrics and leads. There may be other reasons. But the total noise current at the grid terminal of a triode, sometimes referred to as "induced grid noise," is known to be a function of the phase difference between cathode current and anode current; measured values of noise factor show that the uncorrelated component of grid noise current is usually proportional to this total current. This statement may seem like an elaborate method of stating that the noise factor of a tube is a function of the transit time, but it may be worthwhile to point out that phase delay caused by transit time should not of itself produce uncorrelated noise components at input and output ports; therefore, it may be possible to improve tube noise factors substantially by discovering and minimizing other sources of noise. However, other things being equal, a tube designed with reduced grid-wire size and reduced electrode spacings does have reduced noise factor.

Design of Intermediate-Frequency Amplifier Tubes

M. B. Knight

Harrison

The intermediate-frequency amplifier stages of a superheterodyne receiver may be correctly considered as radio-frequency amplifier stages, but intermediate-frequency (if) amplifiers represent a special case because of the nature of the application. Because an if-amplifier stage uses fixed-tuned circuits, it is usually possible to design for higher gain than is feasible for an rf stage with variable tuning. For the same reason, it is usually easier to neutralize the if stage. Shot noise is seldom an important consideration because there is usually signal amplification preceding the if stage. Intermediate-frequency amplifier stages may be required to operate at frequencies ranging from tens of kilocycles per second to hundreds of megacycles; the bandwith may range from hundreds of cycles to tens of megacycles. The major performance considerations of if stages are gain, bandwidth, stability, distortion, and the effects of change in gain on the other three items. The tube design considerations which influence these amplifier performance considerations are discussed in this chapter.

The characteristics of tubes important to their performance as amplifiers are conveniently described in terms of four-terminal (two-port) network parameters: short-circuit input admittance (input conductance and capacitance), short-circuit output admittance (output capacitance and plate resistance), forward transfer admittance (usually grid-to-plate transconductance), and feedback admittance (usually grid-to-plate capacitance). Practical tube design requires that some compromise be made among these parameters. The compromise to be made will depend on the application. For example, the 455-kilocycle intermediate frequency in radio receivers permits design of high-impedance tuned circuits with capacitances of the order of 100 micromicrofarads. As a result, tube input and output capacitances are of little concern. Input resistance is usually very high at this frequency, plate resistance must be made high in order that the high circuit impedance can be conveniently utilized. Even with moderate transconductance values, the gain can be so high that stability becomes a limitation; therefore, grid-to-plate capacitance must be as low as possible.

By way of contrast, the impedances of the tuned circuits in a 45-megacycle if amplifier for a television receiver are usually limited by the tube input and output capacitances. The capacitance limitations, together with the rather large bandwidth requirements of a television receiver, lead to a low-circuit impedances. High gain requires low input capacitance, low output capacitance, and high transconductance; although low feedback capacitance is important, it may be compromised somewhat in favor of low input and output capacitances. Input loading is significant at a frequency of 45 megacycles and should be minimized, but the plate resistance of pentodes is usually so much higher than the practical load impedance that plate resistance is often sacrificed in favor of reduced output capacitance. (See section on OUTPUT ADMITTANCE.)

Pentodes are almost always preferred to triodes for if amplifiers because the inherently high plate resistance and low feedback capacitance of a pentode permits higher gain than feasible with a triode. Triodes are used occasionally when noise is an important consideration as, for example, in the case of a first if amplifier following a crystal mixer. Design considerations for such triodes are the same as for rf amplifiers and, therefore, the remainder of this chapter will discuss only the use of tetrodes and pentodes.

GRID-TO-PLATE TRANSCONDUCTANCE

The best-known factor affecting the gain of an amplifier stage is the grid-to-plate transconductance of the amplifier tube. For a general analysis, this characteristic is considered to be a transfer admittance because transit-time effects cause a phase shift between the grid voltage and the plate current. The phase shift has no direct effect upon amplifier gain, it enters in only through consideration of feedback and then is of practical importance only at frequencies of tens of megacycles and higher.

Circuit designers usually seek the highest values of transconductance obtainable within tube cost limitations. In high-frequency, broad-band circuits, gain per stage is directly related to transconductance; at lower frequencies, tuned circuits can be designed to yield all the gain permitted by stability considerations with tubes having relatively low values of transconductance, but a higher-transconductance tube and a lower-impedance circuit may be more economical. Tube design procedured for obtaining the desired transconductance are discussed in other articles.

It is desirable to obtain the necessary transconductance with a minimum of plate current and at low plate

and screen-grid voltages to conserve power drain and to keep electrodes at a moderate temperature. Some power output is required, however; thus, a theoretical minimum is set for power input. Practical limitations imposed by tolerable distortion require minimum dc power inputs to be much higher than the theoretical minimum set by efficiency considerations. This limitation is often of importance in the last if amplifier stage and may preclude the use of gain control on that stage.

INPUT ADMITTANCE[1, 2, 3]

Input capacitance may be a major factor in determing how large the impedance of the tuned circuit with the given bandwidth requirements may be. For large gain in high-frequency, wide-band amplifiers, therefore, the input capacitance should be small. Even in circuits using an external tuning capacitance, a small tube-input capacitance is desired so that it will have a minimum effect on the tuning and, therefore, frequency stability of the circuit.

Input capacitance is the capacitance between the grid and all other electrodes except the plate. Capacitance between the grid and the cathode is usually the largest part of the total; capacitance between the grid and the screen-grid makes up most of the remainder. Capacitances of a cold tube may be calculated from electrode areas and spacings, although analytical difficulties arise in precise calculation, just as they do in calculation of space current, when the geometry departs from planar or cylindrical shapes.* Best tube design results when the electrode and spacing dimensions are as small as possible. Simple change of electrode areas (as in parallel operation of tubes) is of no advantage because transconductance and capacitance change in proportion to areas. If the spacings are also changed, however, minimum capacitance for a given transconductance is obtained with small electrodes. In the parallel-plane case, for example, capacitance is inversely proportional to the spacing while transconductance is inversely proportional to approximately the square of the control-grid-to-cathode spacing. Practical restrictions limit the degree to which the structure can be scaled down, however, and the most severe restrictions are usually a consequence of the small grid wire diameter needed.

The control-grid-to-cathode capacitance of a conducting tube increases over the cold value because of the influence of space charge. This increase is about 2 micromicrofarads in tubes of the sort normally used for if amplifiers.

Input conductance, or "input loading," becomes of importance at high frequencies where it influences maximum power gain. It results mainly from two components, both proportional to the square of frequency. One source of loading is a function of transit time. A phase shift between the applied rf voltage and the motion of the electrons, principally in the control-grid-to-cathode region, causes the electrons to absorb energy and, therefore, an apparent input conductance results. The second major source of loading is a result of the inductance of the cathode lead. The rf cathode current develops a voltage that has its phase shifted 90 degrees at the cathode. The rf current through the control-grid-to-cathode capacitance is shifted in phase so that there is a component of current in phase with the applied voltage. As a result, there is an apparent input conductance.

Inductance in series with the screen affects input loading in a manner similar to that of the cathode lead inductance, but the phase shifts are such that the apparent input conductance is negative. In some circuits, inductance is intentionally added to the screen-grid circuit to reduce the net input conductance. Too much inductance, however, leads to amplifier instability.

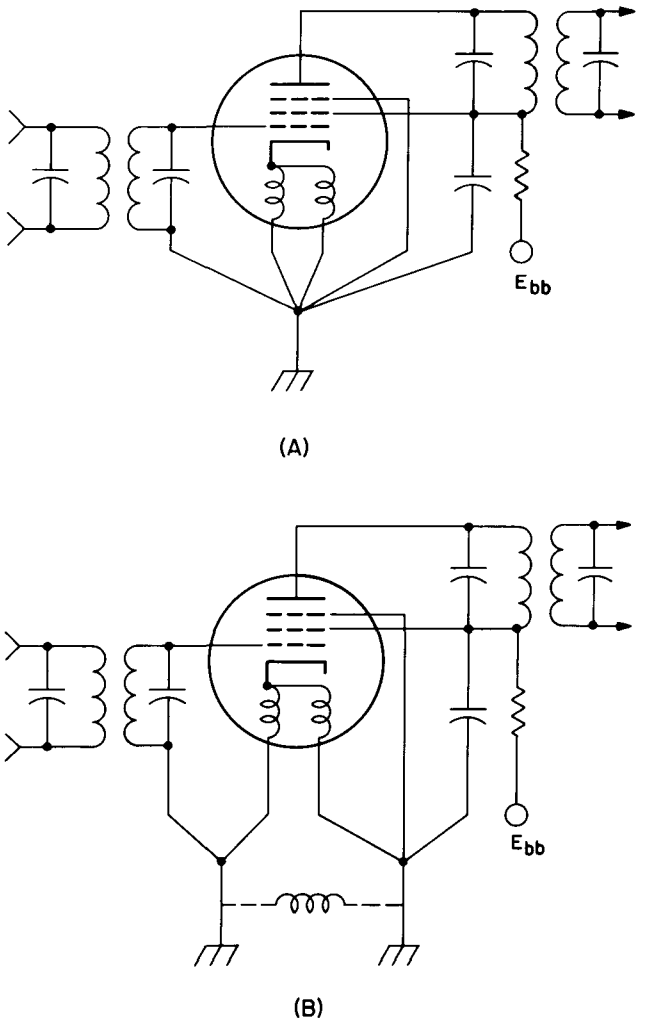


Figure 1. Use of Double Cathode Leads in IF Amplifiers: (A) parallel connection of cathode leads gives lower inductance than single lead; (B) proper connection to different chassis points gives some isolation of input and output currents so that cathode-lead inductance is not common to input and output circuits.

*Approximations can be made using basic formulas for the simple geometrical cases. The simple formulas may be found on pages 135-136 of Ref. 4.

For minimum transit-time loading, small electrode structures are called for. In fact, the considerations are essentially those for obtaining the desired transconductance with minimum capacitance; the quotient of the transconductance divided by the control-grid-to-cathode capacitance is a good figure-of-merit indicator for the transit-time component of input conductance. It is to be noted that the ratio of transconductance to plate current is in itself of no significance to input loading because favorable values are easily obtained with relatively large spacings and large cathode areas.

Cathode-lead inductance in conventional tubes having wire leads, is most easily minimized by the use of multiple leads. In addition to the inductance reduction obtained with parallel leads, double cathode leads permit circuit arrangements in which input and output currents flow in separate leads, minimizing the inductance common to input and output circuits. (See Fig. 1.)

Small components of input loading are introduced by dielectric losses in the stem (usually very small) and series resistance of the leads and cathode coating. The rf losses in the cathode coating depend upon temperature and, in a well-activated cathode, are low at both room temperature and at normal operating temperature. They may, however, be several times higher at intermediate temperatures.

OUTPUT ADMITTANCE

Output capacitance is the capacitance between the plate and all other electrodes except the grid. It comprises mainly the capacitances between the plate and the suppressor or beam confining plates, the screen grid, and the internal or external shield, if used. The plate should be only large enough to capture the available space current and to dissipate the necessary power. The spacing of the plate to the suppressor grid or beam-confining plates is predetermined by design for proper suppresion and high plate resistance. For some applications, however, it may be permissible to reduce output capacitance by increased plate spacing at the expense of optimum suppression. The presence of shields has a large effect on output capacitance, and, inasmuch as shields are generally used to reduce feedback capacitance, output and feedback capacitances are somewhat interdependent.

Output conductance is more commonly known in its reciprocal form, plate resistance. There is no first-order relationship between transconductance and plate resistance in pentode design. Because the maximum available gain is directly proportional to the plate resistance, all reasonable measures should be taken to design for high plate resistance. Many circuits do not realize the maximum available gain because of practical difficulties of matching circuit impedance to the high plate resistance. In such applications, the value of plate resistance may be of secondary importance, but any compromise should be considered carefully.

Plate resistance is determined by three general factors in tube design, the mu-factor of the grids in combination, the effects of secondary emission, and the distribution of electron trajectories[6]. The effective mu-factor of the grids is the predominant factor when the plate voltage is substantially above the screen-grid voltage. This mu-factor may be calculated roughly as the reciprocal of the sum of reciprocals of the mu-factor of each grid to the next electrode. The plate resistance is calculated approximately from the mu-factor and transconductance in the usual way. Actually, this plate resistance is a function of plate voltage because the approximate 3/2-power relationship of plate current to plate voltage becomes apparent at very high plate voltages. The designer usually has the most latitude in adjusting the mu-factor of the screen grid because other factors may control the designs of the other grids. A screen grid with high mu-factor, however, tends to intercept considerable current unless it is made with very fine wire. The best compromise depends upon the intended circuit application.

When the plate voltage is close in value to the screen-grid voltage, the plate resistance may be influenced by secondary emission from plate to screen grid or from screen grid to plate (Fig. 2). Secondary electrons flow from the screen grid to the plate when the plate is above screen-grid potential. Because the secondary-emission current becomes saturated at higher voltages, however, the plate resistance is seriously affected only when the plate is slightly above screen-grid potential. Design of a suppressor grid to prevent these secondary electrons from reaching the plate is not critical because it is necessary to depress the potential between screen grid and plate only slightly below screen-grid potential. For the same reason, it is not difficult to suppress secondary emission from plate to screen grid when the plate voltage is nearly the same as the screen-grid potential. The design becomes more critical at low plate voltages. The potential in the plane of the suppressor grid must be below the plate voltage in order to prevent passage of secondary electrons, but the potential may become low enough, particularly near the suppressor grid wires, to turn back some of the space current. The nonuniform field of the suppressor grid, being at lower potential near the wires and at higher potential in the spaces due to the influence of the plate and screen-grid potentials, has an adverse effect on plate resistance. Best suppression may be obtained with a suppressor grid designed to operate at a positive voltage of about 10 per cent of the screen-grid voltage because the field is then more nearly uniform. Circuit economy, however, usually dictates a zero-potential suppressor grid. The problem is further complicated by the space charge which results from the low electron velocity at low plate

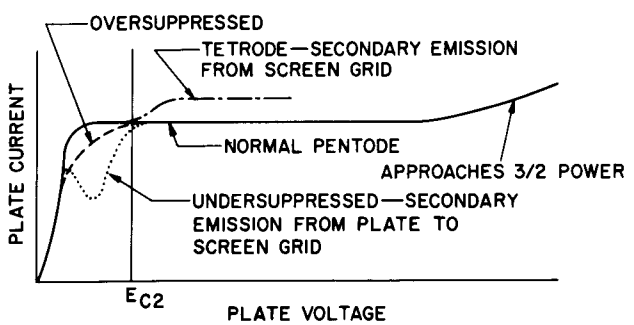


Figure 2. Pentode Plate Characteristic Curve Illustrating the Effects of Several Design Factors on Plate Resistance.

voltages. Thus, optimum suppressor-grid design for low plate currents yields oversuppression (rounded "knee") at higher currents, while optimum design for high currents permits secondary electrons to reach the screen grid at low plate currents.

Distribution of electron trajectories, the third factor influencing plate resistance, also influences suppressor-grid design for low plate voltages as previously discussed. If the electron trajectories were all perpendicular to the plane of the suppressor grid, a uniform-field suppressor grid would give ideal performance. However, most electrons have transverse velocities and there is, therefore, a range of plate voltages where the perpendicular-trajectory electrons will pass and where some of those with angular trajectories will be turned back. Some transverse velocities cannot be avoided because the initial velocities of electrons emitted from the cathode surface start them at various angles. Further deviations are added when the electrons pass near grid wires, i. e. the elctrons are deflected by the non-uniform fields of the grids. Good electron-beam formation, as is obtainable with aligned grids, is the best means to minimize transverse velocities but is seldom practical in if amplifier tubes.

Consideration of the several factors influencing plate resistance reveals an important practical implication: design modifications to improve plate resistance at one plate voltage may even impair the plate resistance at another voltage. For example, increasing the mu-factor of either the screen grid or suppressor grid will increase plate resistance at high plate voltages, but, due to oversuppression or increased transverse velocities of the electrons, may reduce it at low voltages.

FEEDBACK CAPACITANCE

The short-citcuit input and output admittances previously discussed are, in general, not those observed in actual circuits in which the other network terminals are not shorted. Feedback admittance (usually pure susceptance) is always present in actual circuits. Signal applied to the input terminals produces voltage across the load circuit; this voltage causes a current to flow into the input circuit through the feedback capacitance. The phase of this current is determined by the transit angle (susceptance value of the transfer admittance) and by the load admittances; the phase may appear to be that due to a capacitance, inductance, or resistance, including negative values. In a similar way, voltage applied to the output terminals produces a current through the feedback capacitance and, thus, a voltage across the input terminals. This voltage produces output current through the transfer admittance; the effect appears as another component of output admittance.

The various possible phases of the feedback current may lead to interaction of the tuning of input and output circuits, may reduce amplifier gain, or may increase gain (with narrow bandwidth) even to the point of oscillation. The tolerable value of feedback capacitance depends upon the circuit, but the smallest value possible without undue sacrifice of other characteristics is preferred.

Components of feedback capacitance are contributed by the active tube electrodes, by "parasitic" capacitances elsewhere in the tube structure, and by the socket and its associated wiring. Some capacitance between the active parts of the control grid and plate is unavoidable in practical tubes. This component is determined by the electrode areas and by the mu-factors. This component must, therefore, be a compromise with the many other factors which affect the design of the electrodes. It may be indirectly measured by measuring the parasitic capacitances by use of a special tube made with a solid screen grid. The parasitic capacitances may then be subtracted from the total capacitance obtained with a normal tube.

Design for minimum parasitic capacitances is an obscure subject. Calculation is usually not practical because the leads and connectors often have intricate shapes and depart greatly from the extensive surfaces assumed for simple capacitance formulas. The importance of minimum area and maximum spacing of leads and connectors is apparent, but the use of shields requires more study. A combination of theory and experience saves time compared to purely cut-and-try design methods. Capacitance values are determined by electrostatic fields, and the designer can get a good "feel" for shield design through visualizing fields (a by-product of other tube design experience) plus practical experience gained at the console of a capacitance bridge by whittling shields for a tube mount with removable bulb.

An extensive grounded shield between two parts or a shield surrounding one of the parts eliminates the capacitance between the parts. (See Fig. 3.) A shield placed alongside the parts may considerably reduce the capacitance between the parts. A most important factor, but one difficult to handle analytically, is the influence of dielectric materials. Most of the commonly used insulating materials, such as glass, mica, and ceramics, have dielectric constants of about 7. Two wires passing through such a material could just as well be seven times as long or about one-seventh as far apart in a vacuum for the same capacitance. For example, an unshielded glass bulb has an effect similar to that of moving the plate close to the stem leads.

A top mica shield is used to shield the "parasitic" siderods and plate ears which extend through the mica and, especially, to prevent the considerable added capacitance caused by the mica dielectric. These shields are almost obligatory in if amplifier tubes because their use reduces the feedback capacitance by a worthwhile amount without causing much increase in the input and output capacitances. The shield is usually connected to the cathode or suppressor grid but may, instead, be connected to the screen grid so that it can serve as a heat radiator for the screen grid. The area of contact needed between shield and mica may be small if the shield is properly placed and if intimate contact is obtained between the mica and the shield. It is better practice, however, to have the shield cover a large area so that the intimacy of contact is not so critical. In fact, it is common practice to space the shield away from the mica with the aid of "dimples" so that possible surface leakage paths on the mica are longer. Further-

more, this method permits effective outgassing of the shield without excessive heating of the mica.

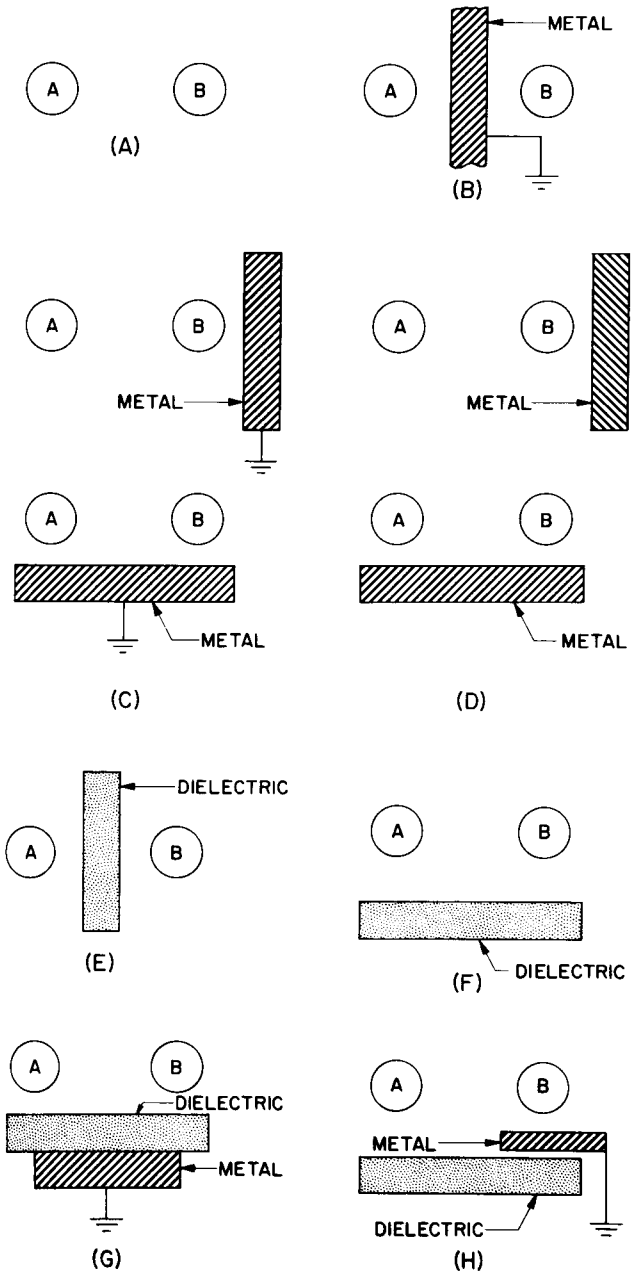


Figure 3. Illustrations of Effects of Metal and Dielectric Materials for the Simple Case of Capacitance Between Parallel Wires: (A) parallel wires, A and B, in vacuum; (B) C_{AB} reduced to zero by complete metallic shield; (C) C_{AB} less than (A), due to grounded shield; (D) C_{AB} greater than (A), due to indirect capacitance caused by "floating metal plate"; (E) C_{AB} greater than (A) due to dielectric material having dielectric constant greater than unity; (F) C_{AB} greater than (A) due to dielectric material; (G) C_{AB} less than (A); (H) C_{AB} less than (F).

The bottom mica shield is used in the same way, and may serve the additional function of providing shielding between the stem leads. One member of the shield may be carried down to the stem close to the control-grid lead. The same member may also touch the glass, preferably with a large-area contact, to reduce the capacitance caused by the glass with its relatively high dielectric constant. Such a shield may be more significant than is apparent from capacitance measurements because conventional sockets do not shield the external surface of the stem as effectively as do the standard-capacitance adapter sockets used for tube measurements. It is often effective to add to the bottom mica shield a member lying between the control-grid lead and the bulb, especially if no internal or external bulb shield is used. This shield member eliminates some of the capacitance between the control-grid lead and the plate via the bulb.

The presence of the glass bulb adds an appreciable component to grid-to-plate capacitance in the manner just described, especially in the absence of a close-fitting bulb shield. The most effective remedy is an internal shield surrounding the cage in order to completely shield the plate from the bulb. It is desirable to have this shield make good contact with the bottom mica or, better yet, to penetrate the mica on the side of the grid lead. The disadvantage of the internal bulb shield is that its use increases the output capacitance. A small shield, even a straight piece of wire, penetrating both micas between the bulb and the plate half nearest the control-grid lead reduces the control-grid-to-plate capacitance almost as much as a full shield, but adds much less output capacitance.

Two inevitable components of feedback capacitance in single-ended tubes are the capacitances in the stem and socket. The significance of these components to the designer working within the restrictions of a given stem design is their effect on the choice of basing arrangements; it is usually not practical to go beyond simple shielding measures for the tube stem or socket. Some approximate values for capacitances in 9-pin miniature button stems and sockets are shown in Table I to facilitate considerations of basing arrangements. Precise values depend upon the conditions of measurement and, if needed, must be measured for the particular

Table I

Capacitances In Nine-pin Miniature Button Stems and Sockets

Pins Measured Between	Direct Capacitance ($\mu\mu$f)			
	9-Pin Stem	9-Pin Stem With Shielding	9-Pin Wafer Socket	9-Pin Molded Socket
3 and 4	0.11	0.1	0.3	0.27
3 and 5	0.014	0.012	0.018	0.02
3 and 6	0.005	0.0025	0.002	0.005
3 and 7	0.002	0.001	0.001	0.002
3 and 8	0.001	0.0005	0.0004	0.001

situation. Capacitance values shown are with all other pins and sockets center-shield grounded; the stem shield referred to is the commonly used disc in contact with the center portion of the button.

In some critical circuits, feedback is caused by mutual inductance between adjacent stem leads. The one troublesome circumstance has been the placement of the suppressor lead adjacent to the control-grid lead. The first confirmed incident was found in a narrow-band 20-megacycle amplifier using type 6AU6. The rf current flowing in the suppressor-grid lead through the plate-to-suppressor-grid capacitance induced a regenerative voltage in the control-grid lead. The feedback effect was found to be equivalent to 0.005 micromicrofarads control-grid-to-plate capacitance at that frequency. Subsequent similar tube designs have interchanged cathode and suppressor-grid leads to avoid this effect.

GAIN CONTROL

Most if amplifier applications require adjustment of gain; it has been found most convenient to adjust gain by adjusting tube transconductance, particularly in automatic-gain-control (agc) systems. All known methods of changing tube transconductance, however, change input and output admittances and change the capability of the tube to handle large signals with acceptably low distortion. At some time in the past, adjustments on each electrode have been tried, including adjustment of filament voltage in tubes having tungsten filaments. Only control-grid-voltage adjustment or suppressor-grid-voltage adjustment satisfies the need for low consumption of power from an agc circuit.

Control-grid-voltage adjustment is used almost exclusively now despite the problems of input admittance variation and distortions. An additional difficulty is that large signals, such as noise impulses or the transmitted pulse in radar, may cause grid-current flow; the resulting voltage developed across the resistance of the agc circuit temporarily disables the amplifier. The remedies for these problems have been adequate to build many successful receivers, even though equipment designers are never fully satisfied with the state of the art.

At first thought, adjustment of suppressor-grid voltage appears to offer attractive features. Control of an electrode not receiving the signal obviates the difficulty caused by rectification of impulse noise and, therefore, high-impedance agc circuits may be used to overcome the lower voltage-sensitivity of the suppressor grid as compared to that of the control grid. Although it would seem that the input admittance should not vary as a result of control with an "outer grid," the variation caused by returned electrons is very serious except in tubes especially designed to direct the space current returned by the suppressor grid so that it does not get back to the control-grid-to-cathode region. The fact that input loading increases as gain is reduced has occasionally been used to compensate for the decrease in input loading observed with control-grid adjustment by using the two control systems simultaneously. Either of these elaborate solutions leaves the problem of a serious variation in the plate resistance of pentodes. Pentagrid structures would preserve high plate resistance, but are not practical because of higher cost and because the additional screen-grid current represents a loss of transconductance and a waste of power. But the crushing blow to suppressor-grid control is that it gives poor results from the viewpoint of distortion. It would seem that the signal-handling capability of the control grid would be unaffected by suppressor-grid control. Even if this were true, however, the signal output capability would fall with the gain, an unsatisfactory situation in most if amplifier applications. The experience of several investigators, this writer included, has been that the tube characteristics develop curious "wiggles" with large reductions in transconductance, and that the distortion is quite bad. Therefore, control-grid voltage adjustment remains the best choice for gain control of if amplifiers.

As plate current and transconductance are reduced with increase in grid bias voltage, the space-charge conditions are altered and, therefore, input conductance and capacitance are reduced. These effects cause changes in the tuning and in the bandwidth of the amplifier. The problem has been handled by a circuit trick to "spoil" the input circuit at low-gain conditions in order to obtain fairly constant input conductance and capacitance. A small, unbypassed resistor is placed in the cathode circuit. When the tube is cut off this resistor is in series with the control-grid-to-cathode capacitance, forming a "lossy" capacitance element. As the tube is allowed to conduct, the rf cathode current produces a voltage across the resistor in proper phase to reduce the rf control-grid-to-cathode voltage and, thus, to reduce the current through the control-grid-to-cathode capacitance. This method is quite successful if the proper value of resistance is used, usually about 50 ohms. Undesirable features are that some gain is sacrificed by degeneration and that the input conductance is made slightly higher.

The change in output admittance with adjustment of grid-bias voltage is usually much less serious. The plate resistance usually increases slightly as current is reduced, but no appreciable detuning occurs. The significance of changes in plate resistance is generally reduced in practical circuits because the load impedance is usually less than the plate resistance.

Feedback elements are usually not changed by this method of gain control, but the effects of feedback are dependent upon transconductance and, therefore, oscillation or serious changes in bandpass characteristics will occur in unstable amplifiers.

DISTORTION

The problem of distortion is emphasized when gain control used. It is usually required that the final amplifier stage provide constant output voltage as the gain is changed; preceding stages, therefore, generally need to handle larger control-grid signals when gain is reduced even though the output may be less. An ideal situation would be to reduce gain without a change in plate current, but this thought is not offered to present an idea—rather to stimulate ideas. The output-level

problem in the final stage is so serious that agc is often omitted or, at least, reduced in that stage lest the dc input become so small that the required output cannot be obtained.

Distortion is produced by curvatures of the tube transfer characteristics. Second order curvature is seldom of concern because it produces only harmonics and sum and difference frequencies of input signals; any of these spurious products of signals within the passband of the amplifier usually fall outside the passband. Indeed, if the plate current were a linear function of control-grid voltage, transconductance would be constant and gain could not be controlled by adjustment of grid-bias voltage. Third, fifth, and higher odd-orders of curvature are manifested in several undesired effects. If three frequencies are present, one type of distortion product present is a new frequency which is the sum of two frequencies minus the third and which may be in the passband. This result may be apparent in another form if the input contains a modulated carrier and an unmodulated carrier; the originally unmodulated carrier is modulated by the sidebands of the other carrier. Both of these effects may be properly called "cross-modulation," although that term has acquired a more specialized connotation in rf amplifiers where the undesired modulation appears on the desired carrier from a strong signal of a frequency well removed from the desired frequency. In order to make a distinction, cross-modulation of signals within the desired passband is sometimes called "intermodulation." A third manifestation of distortion from odd-order curvature is an envelope distortion which changes (usually increases) the apparent modulation percentage and yields harmonic distortion in the demodulated signal. "Modulation rise" is a term frequently applied to this type of distortion.

System performance is best measured by use of one of the effects described earlier. Although various components of the system may compensate distortions of other components, such compensating effects are difficult to control. It is, therefore, desirable to minimize the distortion in each component. Measurement of distortion in a tube is conveniently handled by direct or indirect measurement of curvature of the transfer characteristics without resorting to tests in an if amplifier. Common methods of measurement include mathematical analysis of the transfer curves, measurement of change in transconductance with signal amplitude, and measurement of third-harmonic distortion. These methods will be discussed.

Third-order distortion is proportional to the square of the signal amplitude and to the third derivative of plate current with respect to grid voltage divided by the first derivative. The first derivative is transconductance, so derivatives are more accurate and more easily obtained from direct measurement of transconductance. Because the curve of transconductance versus control-grid-bias voltage must be differentiated twice, it is essential that the original data be very accurate. Good results may be obtained with this method, although it is usually not practical to carry out the differentiation to include the fifth-order distortion. The square root of the reciprocal of the distortion coefficients puts the data in the desirable form of tolerable signal level for a given degree of distortion. There is, however, some error in the regions of the curve having very low distortion. (The distortion coefficient at one point may be zero, suggesting infinite signal-handling capability, but the coefficient obtained by the derivative method becomes invalid for large signals.)

The transconductance change can be measured on a bridge having provision for a large range of input signal amplitudes and a tuned null indication[7]. A preferred procedure is to obtain a transconductance balance with a very small control-grid-signal, to unbalance the bridge a specified amount (usually 3 per cent), and to increase the signal amplitude until the bridge is in balance. The large-signal transconductance is usually higher, though it may be lower in certain regions of the tube characteristics. An advantage of this method is that it gives data directly in the desired form, that is, maximum tolerable signal amplitude for a given amount of distortion.

Another indirect method of evaluating curvature is measurement of third-harmonic distortion (and fifth, seventh, etc. in more refined analyses). The method is straightforward except that the level of distortion being considered is of the order of 1 per cent, so that the requirements for distortion-free signal sources and for measurement accuracy are quite severe.

There are some blind alleys in store for the idealist who wants to design for zero distortion. It is clear from the theory that a square-law plate-current characteristic gives zero distortion because the transconductance curve is then linear and its second derivative is zero. One pitfall is that this linear transconductance curve must remain linear through zero and into negative values, a requirement that implies that the plate current curve must go to a minimum value and then rise with increasing control-grid-bias voltage to complete the parabola. It is inconvenient, at least, to design such tubes although characteristics of this type have been partly achieved by the use of complicated structures. A difficult systems problem arises in applying agc to such a tube; at a control-grid-bias voltage giving near-zero transconductance, a change in bias voltage in either direction may increase gain—the system does not usually recognize the change in the sign of transconductance or in the polarity of the signal. There are also some fundamental restrictions upon tube design. The square-law plate-current characteristic of a single structure cannot be carried down to zero current because the ratio of transconductance to plate current becomes higher than is permitted theoretically by the distribution of initial electron velocities. Avoiding this region by the use of a limited part of the curve which follows a square-law relationship leads necessarily to excessively high plate current.

Plate current tends to be high in any low-distortion tube. When a transconductance curve has been defined, the maximum plate current, the integral of the transconductance curve, is also defined. Thus, any measure to extend cutoff characteristics necessarily increases the maximum plate current. The compromise that has best withstood the test of time is a characteristic having a shape approaching that of an exponential curve; an ex-

ponential curve gives a constant distortion coefficient with change in control-grid bias voltage.

The characteristics of "sharp-cutoff" tubes having uniform electrode geometry throughout the structure have a large rate of change of curvature of the transconductance curve near cutoff. There are numerous methods to extend the cutoff, such as tapered spacings between control grid and cathode or between control grid and screen grid, but any method is subject to the same fundamental relations of transconductance and plate current discussed earlier. The method which has given the best freedom of design and ease of manufacture is the use of a variable-pitch control grid. Some extension of the cutoff characteristic is sometimes obtained with a single gap at the center of an otherwise constant-pitch grid. More stringent requirements lead to refined designs having several gaps of slightly different pitches. The gaps are placed at the center of the grid to minimize the effect of variation in cathode spray length or longitudinal centering of the cathode with respect to the grid. An exceptionally extended cutoff characteristic was obtained by E. W. Herold in unpublished experiments by the use of a very wide gap combined with a cathode in which some of the area under the gap had been made inactive by removal of the cathode coating or by covering it with a band of nickel. These tubes could tolerate very large signal amplitudes but required excessively high agc voltage, a limitation inherent in remote-cutoff tubes.

A circuit technique available for extending the cutoff characteristic is the use of a relatively high screen-grid supply voltage and a series dropping resistor. As the control-grid bias voltage of the tube is increased, the screen-grid voltage rises, and counteracts the change in bias voltage. The resultant condition of high screen-grid voltage and high control-grid-bias voltage reduces the effectiveness of grid control by producing a very nonuniform field in the region of the control grid. The reduction in distortion at low transconductance values is usually found to be in approximate inverse proportion to the screen-grid supply voltage; the tolerable input signal is, therefore, proportional approximately to the square root of the screen-grid supply voltage. This improvement is obtained without compromise in the transconductance or the plate current at high-gain operating conditions. The other methods described improve cross-modulation characteristics only at the expense of less gain or more plate and screen-grid dissipation.

REFERENCES

1. Ferris, W. R., "The Input Resistance of Vacuum Tubes at Ultra High Frequencies," Proc. IRE, Jan. 1936
2. North, D. O., "The Analysis of the Effects of Space Charge on Grid Impedance," Proc. IRE, Jan. 1936
3. Code D1
4. Langford-Smith, F., (Ed.), Radiotron Designer's Handbook, Radio Corporation of America, Harrison, 1953
5. Thompson, J. J., "Measurement and Effect of Cathode-Coating Impedance at Ultra High Frequencies," Proc. of 1958 Tube Techniques Conference
6. Jonker, J. L. H., "Internal Resistance of a Pentode," Phillips Research Reports, Vol. 6, pp. 1-13, 1951
7. Herold, E. W., "Simple Methods for Checking RF Distortion or Cross-Modulation of Pentode Amplifier Tubes," Electronics, April 1940

Design of Converter and Mixer Tubes

T. J. Henry

Harrison

Almost all radio and television receivers now produced operate on the superheterodyne principle. In this system, the incoming rf signal is heterodyned, or shifted in frequency, to a new intermediate frequency which is then fed into an efficient fixed-frequency amplifier of high selectivity and fixed bandwidth. The heterodyne process is carried out by beating the incoming rf signal against a locally generated oscillation signal; the set is tuned by simultaneously varying the constants of the rf and local-oscillator circuits in such a way that the difference frequency remains constant. The intermediate-frequency signal output will, ideally, carry the same modulation components that were present on the rf signal, although in practice some distortion components may arise.

The stage in which the local-oscillator signal is generated and the heterodyning action or frequency conversion takes place is called the converter stage. It has also been called the first-detector, modulator, or heterodyne stage.

Frequency conversion can take place only in a circuit containing a nonlinear element (i. e., nonlinear for output current or voltage vs. input current or voltage), because linear circuits cannot produce any new frequencies. When two voltages of different frequencies f_O and f_S are applied to any nonlinear circuit, a number of resultant new frequencies will be present in the output. Among these will, in general, be the difference frequency $f_i = (f_O - f_S)$ or $(f_S - f_O)$. If the output impedance of the nonlinear circuit is large for f_i and negligible for the other frequencies present, the output voltage will be at the intermediate frequency f_i only. In most receivers, the oscillator frequency f_O is higher than the signal frequency f_S.

Converters may also be operated in harmonic modes where $f_i = (nf_O - f_S)$ or $(f_S - nf_O)$, and n is any integer. In certain converter circuits, the desired output voltage of the stage is the sum $(f_O + f_S)$ instead of the difference of the frequencies. These types of converters, however, are used today in only a few special cases, and need not be of concern here.

When an electron tube (or other electron device) acts as the nonlinear element in the converter stage, and is supplied with an oscillator voltage from an external source, it is called the mixer tube (or device). A tube that performs the functions of both oscillator and mixer is called a converter tube.

In electron-tube mixers, the oscillator and signal voltages may be applied to the same input electrode, or to two separate input electrodes, both acting on the same electron stream. It is generally important to operate the converter under conditions in which the output-signal amplitude is substantially linear with respect to the input-signal amplitude, in order to keep the modulation distortion at a negligible level. Such linearity is closely approached when the oscillator-voltage level is many times greater than the signal-voltage level. This condition of operation is the one usually employed in converter circuits.

The literature contains numerous articles on the design of converter stages, and on the design of special converter and mixer tubes[1]. This chapter will outline some of the problems of designing a converter stage, consider the advantages and disadvantages of various tube designs for specific mixer services, and discuss some of the design problems of the electronics of the converter tube structure.

The interested design engineer will find it to his advantage to read Chapter 25, Section 1, in the Radiotron Designer's Handbook[2].

CHARACTERISTICS OF CONVERTER AND MIXER TUBES

The basic characteristic of the converter or mixer tube is the conversion transconductance g_c defined as:[3] "The quotient of: (1) the magnitude of the desired output-frequency component of current by (2) the magnitude of the input-frequency (signal) component of voltage when the impedance of the output external termination is negligible for all frequencies which may affect the result."

This definition ordinarily assumes that the rf input-frequency voltage is small. The conversion transconductance is a function of the oscillator-voltage amplitude as well as of the dc voltages on the tube electrodes.

Conversion transconductance may be used to determine the small-signal conversion gain (neglecting feedback effects) in the same manner as grid-plate transconductance (g_m) is used for rf pentodes.

$$\text{Conversion Gain} = g_c \frac{R_L}{1 + (R_L/r_p)}$$

where R_L is the load resistance and r_p is the effective plate resistance of the mixer tube with oscillator voltage

applied. The magnitude of r_p varies with the strength of oscillation but under any given voltage conditions may be defined as the inverse of the rate of change of dc plate current with plate voltage.

The conversion transconductance of any mixer tube may be computed from curves of signal-electrode-to-anode transconductance as a function of the oscillator-voltage amplitude. This method assumes that the signal voltage is small and that the oscillator voltage is large, reasonable assumptions at low signal levels. By use of a Fourier series analysis, the conversion transconductance g_c can be shown to be[2]:

$$g_c = \frac{1}{2\pi} \int_{-\pi}^{\pi} g_m \cos\omega_0 t \, d(\omega_0 t)$$

where g_m is the transconductance from the signal electrode to the anode, and the oscillator voltage e_0 varies at the rate $E_m \cos\omega_0 t$ around a dc bias voltage, E_0. See Fig. 1.

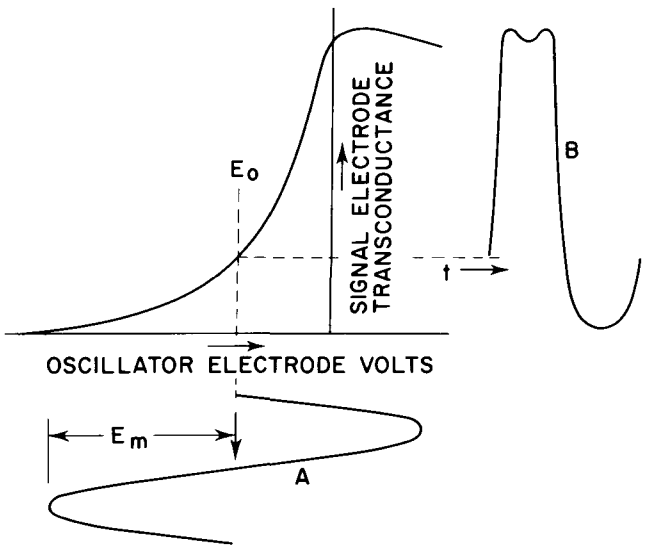


Figure 1. Signal-Electrode Transconductance Versus Oscillator Electrode Voltage for a Typical Mixer Tube (applied oscillator voltage is shown at A; B is the resulting time variation of transconductance)

An examination of the above equation shows that the optimum g_c is obtained when g_m is zero during the negative excursions of $\cos\omega_0 t$, and is at a maximum positive value $g_{m(max)}$ for all positive values of $\cos\omega_0 t$. Under these conditions,

$$g_c = \frac{g_{m(max)}}{\pi} \quad *$$

Once a curve of signal-electrode transconductance vs. oscillator voltage has been obtained, the integral may be evaluated approximately by the trapezoidal rule[4].

*Higher values of g_c are possible in a tube in which a phase reversal of g_m takes place when $\cos\omega_0 t$ passes through zero so that the product $(g_m \cos\omega_0 t)$ is always positive. In this case the theoretical maximum $g_c = 2g_{m(max)}/\pi$. Successful mixer tubes have been constructed in which a phase-reversal of g_m has been obtained by means of beam deflection control of the electron stream[5].

A tube with a characteristic similar to the one shown in Fig. 1 will have a value of conversion transconductance of approximately $0.28\, g_{m(max)}$ when the oscillator signal and the bias voltage are optimized. For many purposes, it is sufficiently accurate to assume that the optimum value of conversion transconductance can be taken to be $g_{m(max)}/4$.

As examples of the results obtained with the actual mixer and converter tubes, Figs. 2 and 3 show curves on two popular types. Fig. 2 shows the conversion transconductance of the pentode section of a 6CG8 tube as a function of oscillator voltage. In this curve, the control-grid bias voltage is fixed and the conversion transconductance increases rapidly with oscillator voltage and then falls off somewhat less sharply. The rate of variation of conversion transconductance with oscillator voltage will be considerably reduced when a grid-leak-bias is used. Fig. 3 shows the conversion transconductance from the grid No. 3 to the anode of a 6BE6 pentagrid converter as a function of current to grid No. 1, oscillator grid, when the tube is self-excited.

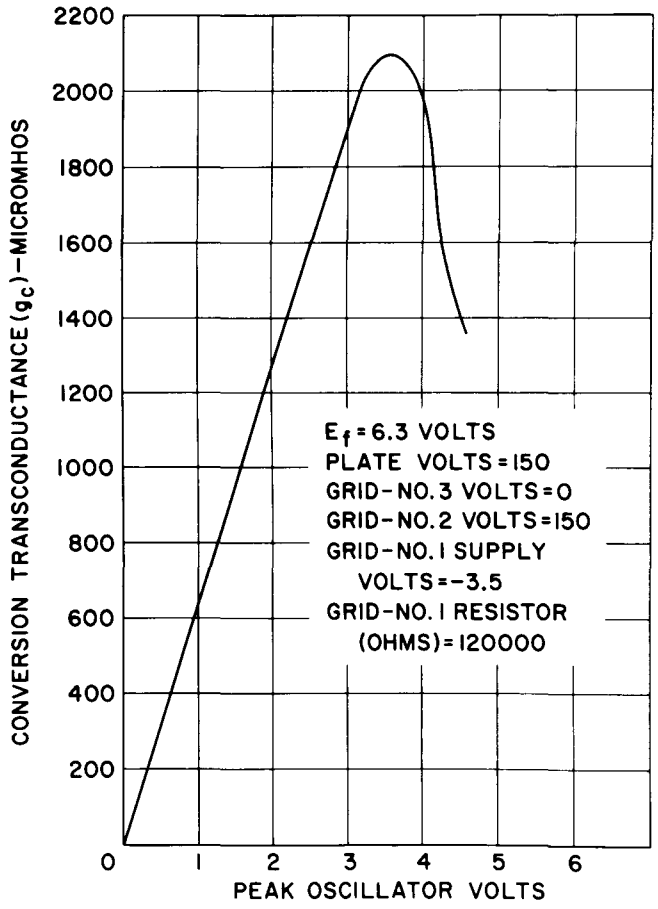


Figure 2. Operation Characteristic of the 6C8G (Pentode Unit) with Separate Oscillator Excitation

This method of analysis for conversion transconductance is a general one and is valid whether the oscillator and signal voltages are applied to the same or different electrodes. The method also applies to computation of the conversion conductance of diode mixers.

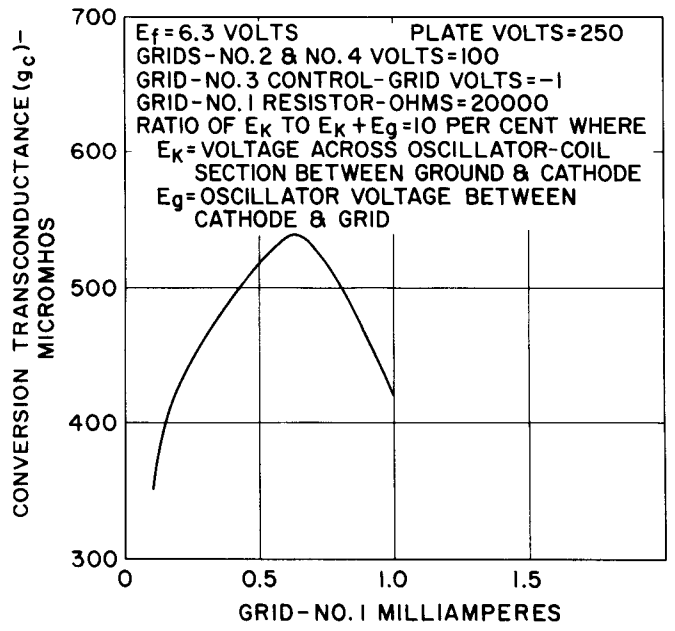


Figure 3. Operation Characteristic of the 6BE6 Converter Tube with Self-Excitation

Many converter stages contain a Hartley oscillator circuit in which the oscillator-frequency voltage is present on the cathode (Fig. 4). This arrangement changes the voltages between the cathode and the other elements over the oscillator cycle in such a way that the transconductance from signal electrode to anode is reduced somewhat at maximum positive oscillator voltage[6]. An accurate analysis for conversion transconductance in this case requires a transconductance curve taken with the potentials on the oscillator-input electrode and cathode varied simultaneously in their proper ratio as determined by the ratio of the cathode turns to the total turns of the oscillator coil which is to be used. It is often sufficiently accurate in the calculation of conversion transconductance to disregard the ac variation of cathode potential and to shift the signal-grid-bias voltage in the negative direction by the peak value of the ac cathode voltage.

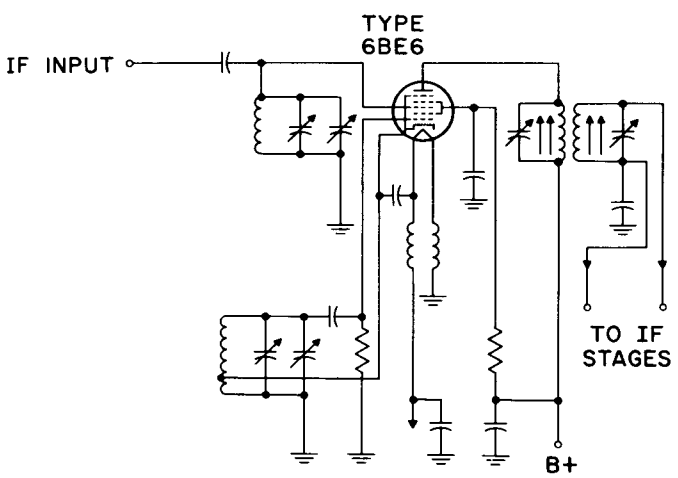


Figure 4. Typical Self-Excited Converter Circuit for 6BE6

It should be noted that a nonlinear mixer action can be obtained with a linear transfer curve so long as the current is cut off during part of the oscillator cycle.

This method of calculating conversion transconductance is helpful in understanding the operating principles of mixer devices and sometimes in analyzing particular designs in the development stage.

For accurate results on quantities of tubes, the conversion transconductance is usually measured by one of several methods that are widely used[7]. The standard method for multigrid types accepted in the industry is described in the same reference as follows:

The definition of conversion transconductance shows that if $f_1=f_2$, the difference frequency will be zero, corresponding to direct current. If the two signals are in phase, the plate current will be higher than with the two signals 180 degrees out of phase. The difference in direct plate current, caused by a reversal of the phase relation of the two signals, divided by twice the peak value of the alternating voltage on the signal grid, is the conversion transconductance. It is assumed that the voltage on the oscillator grid is adjusted to such a value that the current drawn by the grid corresponds to its rated value.

In practice, multigrid converter and mixer tubes may be measured in a circuit such as that shown (in Fig. 5).

With grid resistor R_1 and grid current as rated, a signal voltage E_g is applied to the signal grid 180 degrees out of phase with the voltage applied to the oscillator grid. The bucking circuit is adjusted to give zero reading of the plate-current meter, after which the phase of the signal voltage applied to one of the grids is reversed and the reading of the plate-current meter ΔI_b noted. Conversion transconductance is then

$$g_c = \frac{\Delta I_b}{2\sqrt{2}\,E_g}$$

For precision measurements, the signal voltage applied to the signal grid should be held to as low a value as possible and in no case should cause current to flow in the signal-grid circuit. For most converter tubes, it is convenient to make E_g equal to 0.354 volts rms. The conversion transconductance in micromhos is then equal to ΔI_b in microamperes."

This method is used as a standard, because the results can be duplicated at different locations with a little care in measurement.

Other methods, which involve mixing two frequencies and measuring the output current at the difference frequency as a function of signal-grid voltage, are speedier to use, and are widely used for factory and quality-control tests on converters. Unfortunately, good agreement between test sets, even in the same location, can only be maintained by continual calibration and circuit adjustment. Agreement on readings among various manufacturers is much more difficult to obtain with these methods, and the phase-reversal technique described remains the most useful industry standard.

Conversion transconductance of pentodes and triodes may also be measured, but for production controls and even in development, it is often accurate enough to estimate conversion transconductance from the maximum positive transconductance $g_{m(max)}$ reading under the expected conditions of operation.

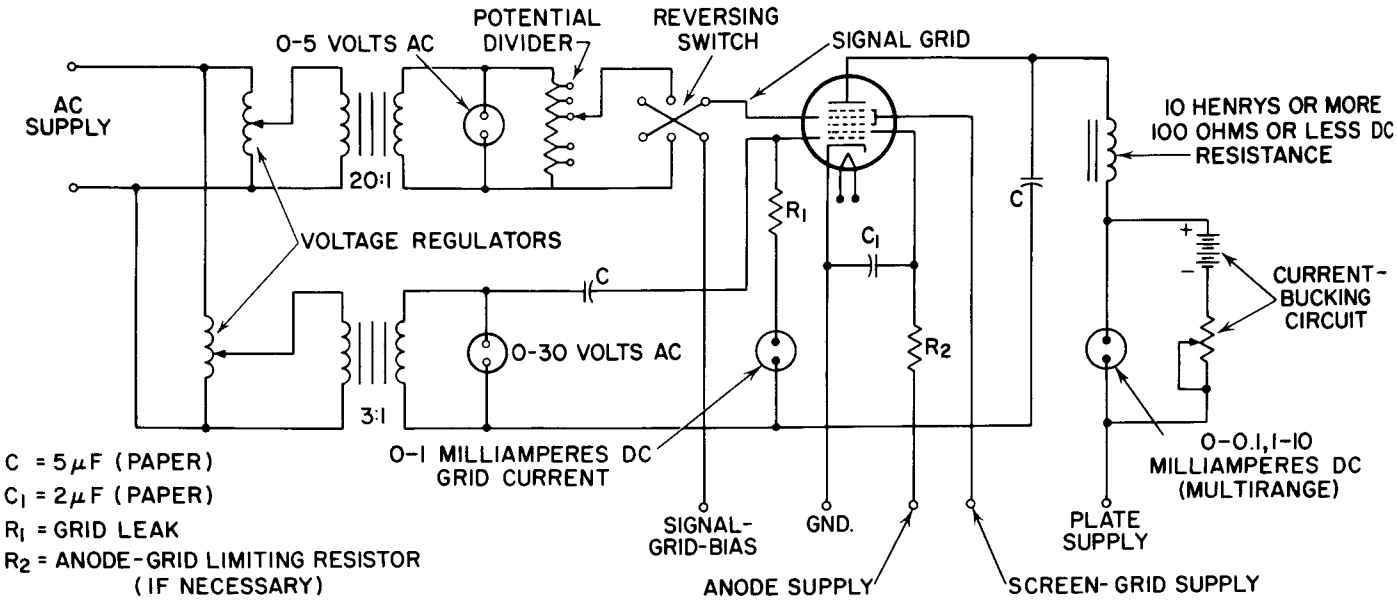


Figure 5. Circuit Arrangement for Measuring Conversion Transconductance

A characteristic curve of measured or computed conversion transconductance (g_c) vs. the control-grid-bias voltage (E_g) is often useful (see Fig. 6). This curve represents performance at one condition of dc element voltages and at one oscillator-voltage amplitude. Such a curve is useful in estimating the amount of cross-modulation that one is likely to encounter if automatic-gain control (agc) bias voltage is applied to the control grid. If the tube has a sharp cut-off so that low values of conversion transconductance cannot be obtained without working very close to cut-off, the use of agc on this tube will not be practical. In remote or semiremote cut-off types, an estimate of the cross-modulation distortion at various bias voltages may be obtained from graphical differentiation of the curve of control-grid bias voltage versus conversion transconductance. The cross-modulation is proportional to

$$\frac{d^2 g_c / dE_g}{g_c} \times E_g^2$$

and it is desirable to keep this factor low over the operating bias-voltage range. More accurate methods of measuring cross-modulation are also available[8].

THE DESIGN OF THE CONVERTER STAGE

The converter stage of a receiver presents a number of special problems to the set designer in addition to other important considerations not unique to this stage.

The converter stage must contain circuits for signal, oscillator, and intermediate frequencies; and in general, must be tunable over a considerable band of signal frequencies. The problems of avoiding interaction of the various circuits throughout the band may become extremely complex. For simplicity of circuit design, it would be desirable for the mixer tube to have three sets of isolated terminals for the three different frequencies and to have the electrostatic coupling through the tube itself minimized. With multigrid-mixer and

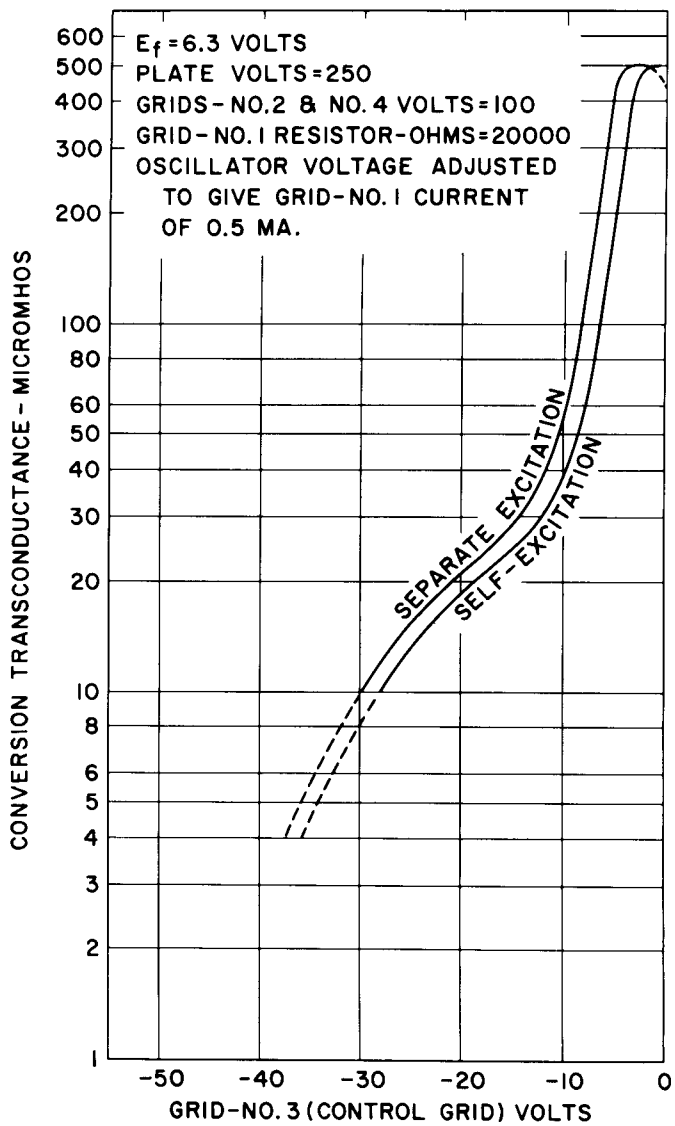


Figure 6. Conversion Transconductance of the 6BE6 Versus Bias Voltage on Grid No. 3 (Control Grid)

converter types, ideal isolation of the three circuits can be closely approached, and therefore, these types are used in most low-frequency services. Above 20 megacycles or so, the faults of such tubes outweigh their advantages. As a result, triode and pentode mixers are generally preferred despite their more complex circuitry.

Some of the areas which present problems to the set designer are listed here. Optimizing the performance in respect to all the important areas can generally not be done and the set designer must choose a compromise design suitable to the particular market demands that the set must meet.

The characteristics of the tubes to be used in the converter stage are an important consideration in circuit design, and the tube design engineer and tube application engineer should be aware of the set designer's problems and requirements in order to provide the best possible tube design for the job.

Automatic Gain Control (AGC)

In many applications, automatic gain control by biasing the signal grid on the converter stage would be of considerable advantage. Use of automatic gain control on the converter requires: (1) that the curve of control grid bias versus conversion transconductance have a remote or semi-remote cutoff and be properly shaped in order to keep cross-modulation low, and (2) that variation of the control grid bias does not materially influence the oscillator frequency and thereby cause detuning as the signal level (and agc bias) is varied.

Tuning

Economic considerations frequently require that receivers be designed to tune over wide bands of frequency (as much as a three-to-one variation in frequency) by varying the tuning capacitance. Over this tuning range it is necessary to have good tracking of the oscillator and signal circuits, and it is desirable to have reasonably uniform gain and oscillator amplitude over the band. This tuning problem must be solved by the set designer with appropriate choice of circuit constants and padding networks. In such circuits, it is not practical to operate the oscillator tube at optimum conditions throughout the band. The tube engineer can contribute to the solution of the problem by providing converter tubes (or oscillator tubes) with sufficient factor of safety in oscillator performance to provide good oscillator excitation throughout the band under relatively inefficient operating conditions.

Interaction of Signal and Oscillator Circuits

When the ratio of oscillator frequency to signal frequency approaches unity, the impedance of the signal circuit to oscillator frequency may be sufficiently high that considerable oscillator voltage may be developed on the signal grid of the converter tube, if coupling exists in the tube. In multiband radio receivers, this action frequently becomes troublesome on the high-frequency bands. The effect is usually either to increase or decrease the relative modulation of plate current at oscillator frequency, and so, to change the conversion transconductance. This is undesirable particularly because the magnitude of the induced voltage usually varies with frequency. The oscillator-frequency voltage developed on the signal grid may also become large enough to exceed the signal-grid bias, and grid current will flow in the signal-grid circuit[9].

Oscillator Strength

In high-frequency applications, when circuit losses become higher, it is important to have a high-gm oscillator tube that can be used with reasonably-priced circuit components that will start readily, and will not drop out of oscillation anywhere in the tuning band, particularly under low line-voltage conditions.

Input Conductance at High Frequencies

At higher frequencies, the input conductance of the signal grid will become high enough to be important in determining the effective resistance of the input circuit, and the gain will decrease because of the increased loading. A similar input loading occurs on amplifier types at high frequencies, but in many mixer designs the effect is much more severe. Input loading is a function of frequency and often makes it difficult to obtain uniform gain throughout a wide frequency range.

Noise

In high-frequency circuits, where tube gain is reduced by input loading and higher circuit losses, the noise generated in the first-stage tube has a marked effect on the signal-to-noise ratio of the receiver[10]. Even when an rf stage is used ahead of the converter stage, mixer noise may be an important consideration if the rf gain is low.

Oscillator Radiation

Local-oscillator radiation from television receivers can be an important source of interference with reception on neighboring television sets, and must be kept low. The rf amplifier stage attenuates any oscillator signal feeding back through the input circuits to the antenna. A mixer without an rf stage ahead of it would not be satisfactory according to present television-receiver standards unless a special type of tube circuit could be devised which produced very little oscillator voltage on the signal-input circuit. It is also desirable to have a mixer that works satisfactorily with low levels of oscillator voltage, in order to minimize radiation directly from the chassis and oscillator circuits.

Frequency Drift During Warm-up of Television Receivers

It is desirable to keep the drift of oscillator frequency on warm-up of a television set below one or two hundred kilocycles in order to avoid the need for constant returning of the set. Warm-up drift can be compensated for by use of temperature-sensitive trimming capacitors, but oscillator types with low inherent drift will reduce the maximum residual frequency deviation[11].

Economic Considerations

In the entertainment market, the balance between cost and performance is very important. It is not enough to make a good technical design of high performance; the product must sell in a sharply competitive market at a cost commensurate with the performance.

TYPES OF MIXERS AND CONVERTERS

Mixer tubes may conveniently be divided into three classes:

1. Tubes in which the oscillator and signal voltages are applied to the same electrode.
2. Tubes in which the oscillator signal is applied to a grid further from the cathode than the electrodes to which rf signal is applied. Such types are called outer-grid-injection mixers.
3. Tubes in which the oscillator signal is applied to a grid closer to the cathode than the electrode to which the rf signal is applied. Such types are called inner-grid-injection mixers.

Figs. 7 through 12 show various types of mixer and converter tubes and their associated circuits.

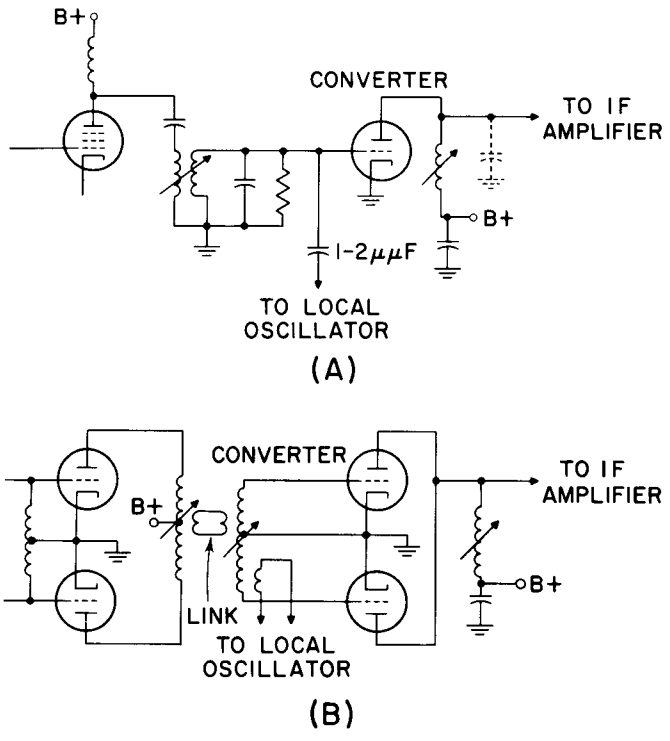


Figure 7. Converter Circuits: (A) triode; (B) balanced twin triode

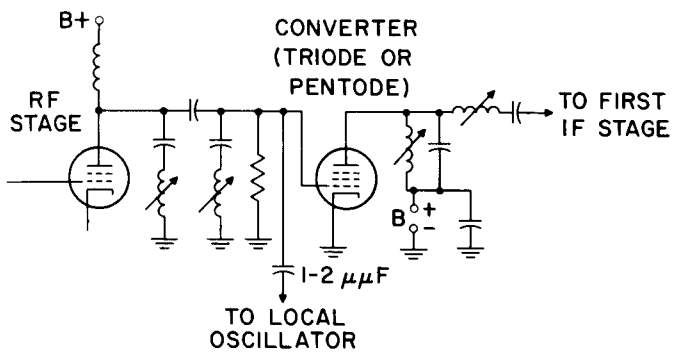


Figure 8. Pentode Converter Circuit

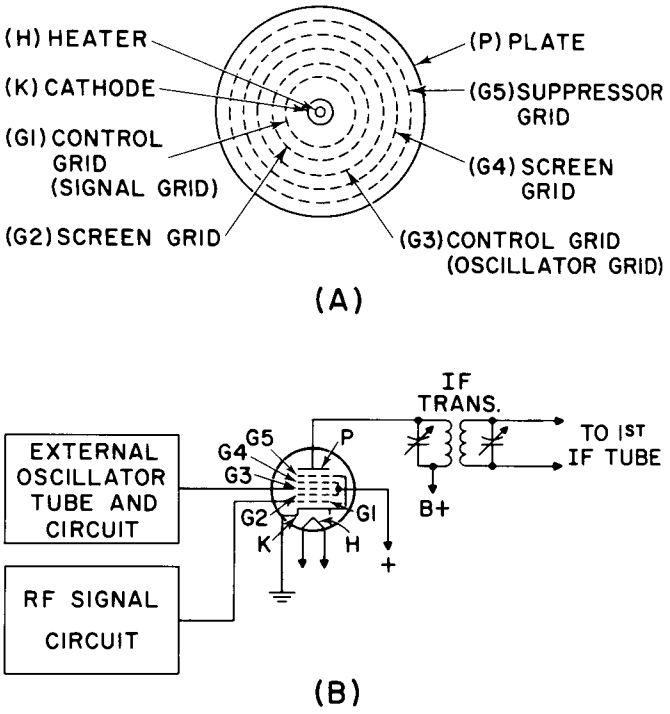


Figure 9. Use of the 6L7: (A) Relative Position of Electrodes; (B) Connection as a Mixer

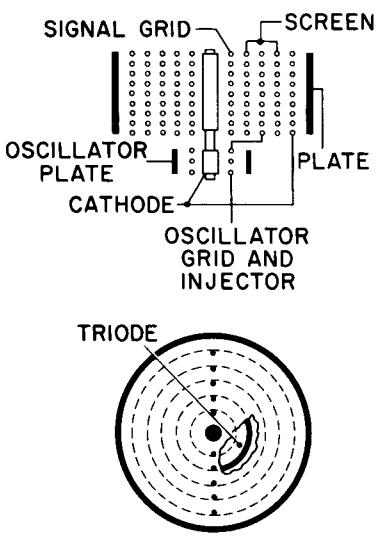


Figure 10. Electrode Arrangement of the Triode-Heptode Type 6J8-G

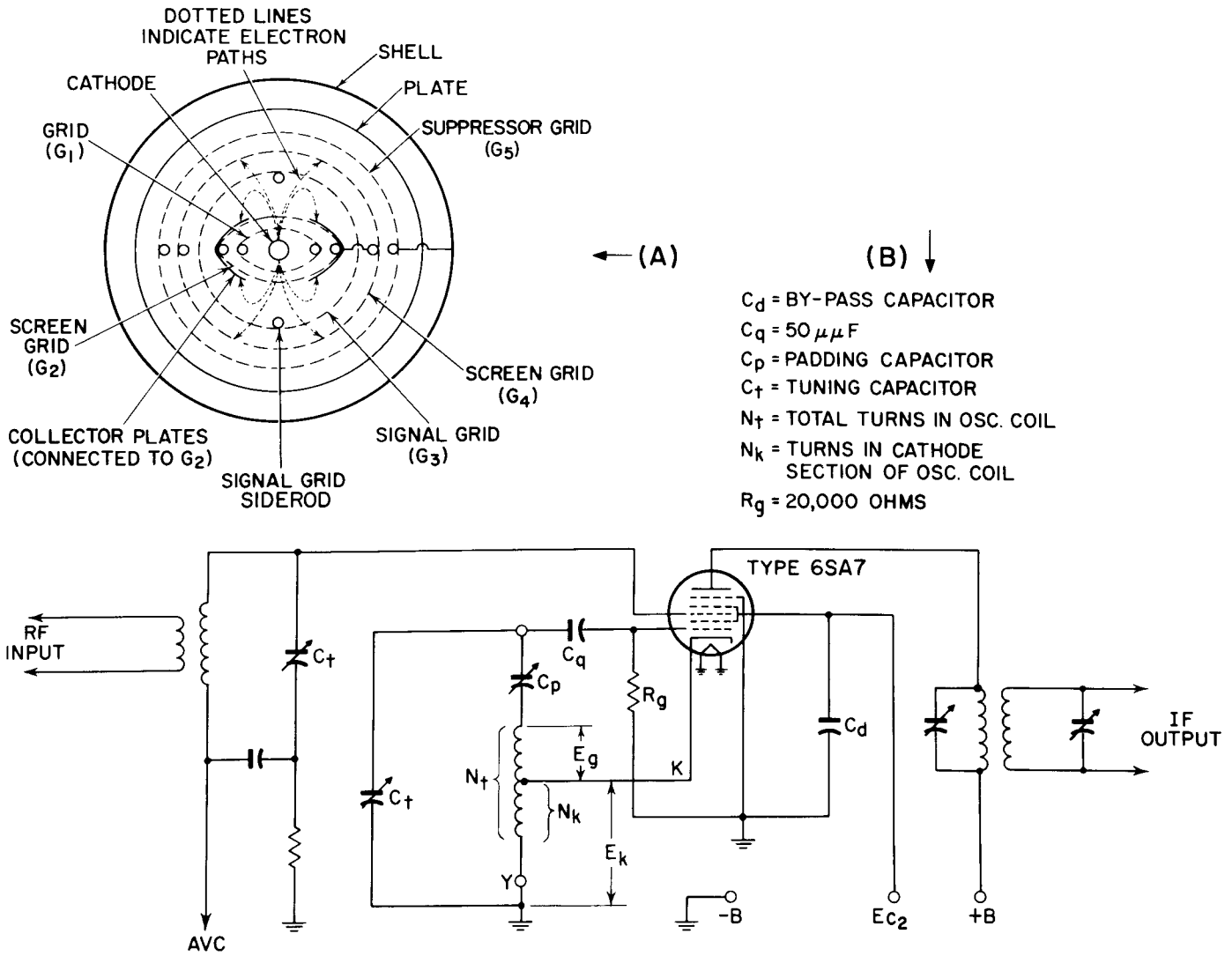


Figure 11. Use of 6SA7: (A) structure; (B) typical self-excited converter circuit

The Diode Mixer

The simplest mixer tube is a diode. Diode mixer tubes always introduce a conversion loss and are very poor for noise factor. They are practically never used.

The semiconductor diode, however, has a wide application as a mixer at frequencies above 1000 megacycles because of its relatively low noise factor at these frequencies. Semiconductor diode mixer stages are used as the front-end stage in almost all uhf television tuners, (i.e., in the 500-900 megacycle range) as being the best compromise design for noise performance and set cost. More recently, several rf-amplifier triodes are available that yield better noise performance than the semiconductor diode mixer as a first-stage device in this frequency range; it is largely a matter of economics as to whether an rf amplifier tube or a semiconductor-diode mixer will be used in the first stage of new designs of uhf television receivers. If an rf tube is used, it will probably be used with a triode tube mixer to obtain better gain and simpler circuitry than would be obtained with a semiconductor diode mixer.

The Triode Mixer

The triode mixer has the best noise figure of any conventional type of mixer tube (excluding traveling-wave tubes or parametric devices); and, when the plate resistance of the tube is high compared with the load impedance, it will give good conversion gain. A disadvantage of the triode is that feedback through the tube loads the input circuit considerably. Neutralization, or the use of specially designed intermediate-frequency circuits with low impedance to signal frequencies may be used to reduce this loading. Ordinarily, the oscillator signal is injected into the grid where oscillator power requirements are less. The use of cathode injection will isolate the signal circuit a little more than grid injection does, but the signal circuit is loaded down by the feedback voltage across the impedance in the cathode circuit[12]. Many of the early television sets used triode mixers or push-pull triode mixers, but today the use of pentodes is more usual.

It is also possible to use a triode (or a pentode) as an autodyne converter, in which the rf signal is applied to

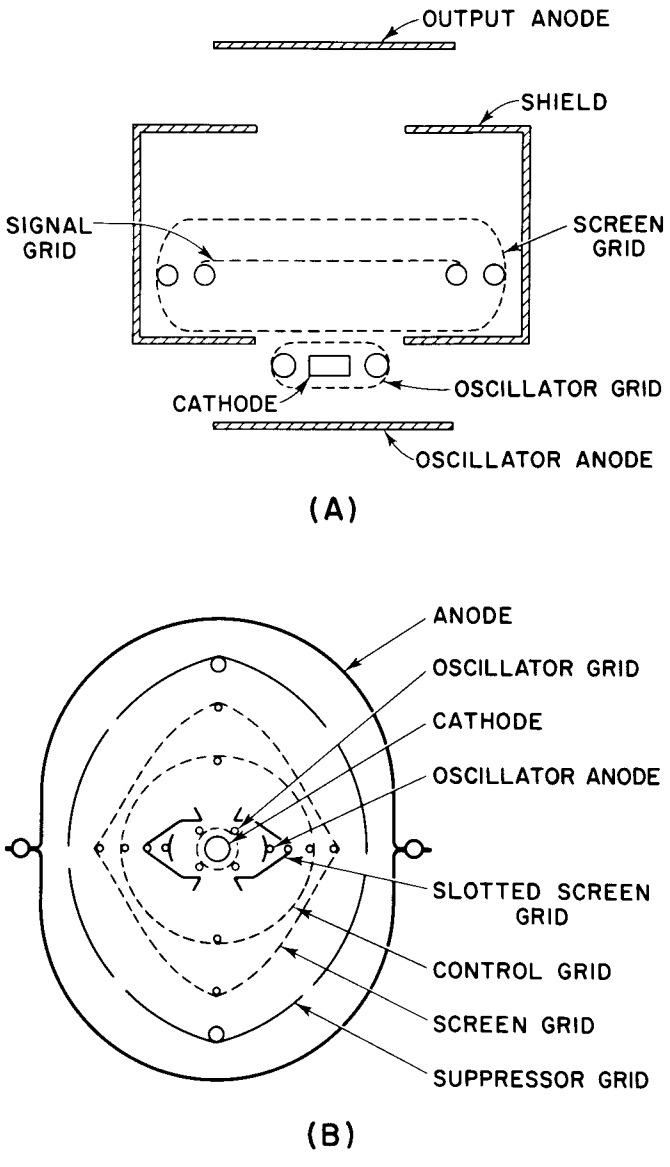


Figure 12. Beam-Type Converter Tubes: (A) cross-sectional view of the 6K8 converter tube; (B) cross-sectional view of the four-beam octode

the grid of an oscillator tube and no separate excitation is needed. Such converters have been made to operate satisfactorily, but are probably more tricky to design and maintain over a tuning range than more conventional circuits.

The Pentode Mixer

The pentode mixer has a somewhat higher noise figure than an equivalent triode with the same front end, but the lower feedback capacitance (Cg_{1-p}) makes the loading effects less critical. In the pentode, the oscillator signal is usually injected in the signal grid, but may be injected in the cathode circuit with the same type of additional loading as on the triode. Oscillator voltage may be injected into the grid No. 3 circuit but this method offers no particular advantage.

The pentode mixer, in general, can be made with good transconductance and high plate resistance so that good conversion gain is realized. The noise factor is somewhat higher than that of the triode, but noise factor becomes fairly unimportant when a high-gain rf amplifier stage is used ahead of the mixer.

Because of its reasonably good noise and gain performance, and its relative freedom from feedback effects, the pentode mixer is used in most present-day television sets, particularly combined with a triode oscillator.

It would be possible to use a remote-cutoff pentode as a mixer and to apply automatic gain control to it. However, the remote-cutoff pentode would require a much higher oscillator voltage to drive it to optimum conversion transconductance values, and oscillator radiation problems would be increased considerably. The noise-factor of a variable-mu pentode would also be somewhat higher. Mixer pentodes currently used in television sets are of the sharp-cutoff type, and do not ordinarily have automatic-gain-control bias applied to them.

The Outer-Grid-Injection Mixer

Examples of the outer-grid-injection mixer are the 6L7 mixer heptode and the 6J8G triode-heptode. In both of these types the grid No. 1 is the signal grid, grid Nos. 2 and 4 are screen grids operating at a positive potential, grid No. 3 is the oscillator injection grid, and grid No. 5 is a suppressor grid tied to the cathode. In the 6J8G converter, the heptode mixer is combined with a triode oscillator section in the same bulb.

The outer-grid-injection mixer has the transconductance largely controlled by the design of the front-end of the structure and can readily be designed to give fairly high values of conversion transconductance. Grid No. 1 may be made with a variable mu for low cross-modulation with automatic gain control. Because the oscillator section performance is not affected materially by grid No. 1 bias voltage, there is little of the tendency toward frequency shift with automatic gain control that is found in some inner-grid-injection converters.

The outer-grid-injection tube, in general, gives less trouble from coupling between signal and oscillator circuit at high frequencies than inner-grid-injection types. The input conductance to the signal grid is positive and may be considerably higher than when the same tube is used as an amplifier. There may also be a dc grid current to the negative signal grid when oscillator frequencies are higher than 20 megacycles or so.

The negative grid current and the very high input conductance of outer-grid-injection types are both due to the oscillator grid returning electrons which pass again into the signal grid region; both effects can be reduced[1] by the use of auxiliary electrodes to reduce the number of electrons that get back through the screen grid (grid No. 2). These auxiliary electrodes may conveniently be attached to grid No. 2.

A well-designed outer-grid-injection mixer will give good performance at low and medium high frequencies.

At higher frequencies, as in television applications, noise and transit-time effects make this mixer type less suitable than triodes or pentodes. The main disadvantage of outer-grid-injection mixing as compared to inner-grid-injection is that a separate oscillator section is required. In order to get a high-transconductance triode for strong oscillation at high frequencies, a large cathode area is required for the triode. As a result, a separate oscillator tube or a tube design with a large cathode and high heater power is necessary. Because the inner-grid-injection converters use the same electron stream for oscillation as well as for mixing, they conserve heater power and make possible a less expensive converter stage with adequate performance.

In the USA, the outer-grid-injection mixer or converter is not used in present-day radio designs. In Europe, combination tubes with a triode oscillator section and a outer-grid-injection mixer heptode section are still popular with radio set designers.

The Inner-Grid-Injection Converter

Commercially, inner-grid-injection tubes are always converters in which the oscillator and mixer functions are combined in one envelope and frequently use the same electron stream for both functions.

The pentagrid converter, as widely used in the USA, has the structure shown in Fig. 11A and uses a typical circuit such as that shown in Fig. 11B. Grid No. 1 serves as the oscillator grid, grid Nos. 2 and 4 are screen grids, grid No. 3 is the signal grid, and grid No. 5 is a suppressor grid. By use of the hartley oscillator circuit, the screen-grid and plate currents together serve as the oscillator anode current and, therefore, strong oscillations are readily obtained. Because the cathode voltage is swinging, a judicious adjustment of the oscillator voltage on the cathode is necessary to avoid degenerating the conversion transconductance and swinging the cathode to a voltage more negative than that on the signal grid[2].

At amplitude-modulation broadcast frequencies, this type of converter tube gives quite satisfactory performance, and is used because of the relatively good circuit isolation. At higher frequencies, in the 6 to 18 megacycle band, internal coupling and admittance effects become important and complicate circuit design.

Because of the two screen grids and the type of operation, the screen-grid current in the pentagrid converter is a large part of the total cathode current (about 60 per cent as compared to 25 per cent in a pentode mixer); and the noise factor is high and the conversion gain is relatively lower than that of a good pentode mixer with the same cathode current. In the amplitude-modulation radio broadcast band, the tube noise in the receiver is small compared to the noise from the circuit impedances and adequate gain is not hard to obtain. The pentagrid converter is the most convenient and economical tube for this service.

The signal grid, grid No. 3, is usually wound to give a remote-cutoff characteristic; automatic gain control can be obtained by changing the bias voltage on grid No. 3 without materially influencing the performance of the other circuits.

In the pentagrid converter as in the outer-grid-injection mixer, it is undesirable to have electrons returning through the screen grid (grid No. 2) into the region between grid No. 1 and the cathode. The number of such electrons will vary with the bias voltage due to automatic gain control on grid No. 3; more electrons will be returned toward the screen grid at high bias voltage on grid No. 3. If these returning electrons reach the grid No. 1 region, they will change the space charge in that region and affect the admittance of grid No. 1. These changes will cause oscillator frequency to shift as a function of rf-signal strength and will make the tuning difficult[13].

COUPLING EFFECTS

With an inner-grid-injection type of mixer, the negative space charge in the vicinity of the control (signal) grid increases as the oscillator grid becomes more positive. This increase in negative space charge tends to drive electrons out through the control-grid circuit. The capacitance between oscillator grid and control grid has the opposite effect, that is, a positive increment of voltage on the oscillator grid will tend to induce a negative charge on the control grid and draw electrons in through the control-grid circuit. If, as is customary, the oscillator frequency is higher than the signal frequency, the input circuit will offer capacitive reactance to the oscillator frequency, and the variation in space charge will produce a voltage out of phase with the oscillator voltage. This electron coupling acts as a negative capacitance from oscillator grid to signal grid. The effect becomes more important at high frequencies because, if the intermediate frequency is kept constant, the signal and oscillator frequencies will have a lower ratio and, therefore, the input-circuit impedance will become appreciable enough at oscillator frequency to develop higher voltages at oscillator frequency across the signal circuit. With higher oscillator frequency than signal frequency, this induced voltage of oscillator frequency on the signal grid tends to reduce the conversion transconductance. With lower oscillator frequency than signal frequency, the induced voltage tends to increase conversion transconductance.

The negative capacitance due to space-charge coupling can generally be neutralized for any particular frequency by connecting an external capacitance from oscillator grid to signal grid, but at very high frequencies, where transit time becomes a factor, a resistance may be required in series with the capacitance. This method of neutralization is not entirely satisfactory. The use of a neutralizing capacitance increases frequency drift in the 6K8.

Outer-grid-injection tubes also show some space-charge coupling due to electrons returning from the oscillator grid to the signal-grid region, but the effect is slight. In the inner-grid-injection type, space-charge coupling is inherently present because the space charge near the signal grid must vary at oscillator frequency.

NOISE IN MIXER TUBES

An idea of the relative amounts of fluctuation noise produced in the various types of mixers may be obtained from a comparison of four fictitious mixers as calculated by Herold[10].

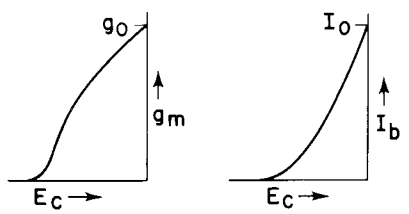
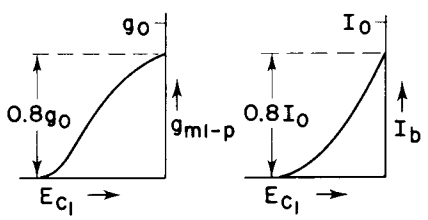
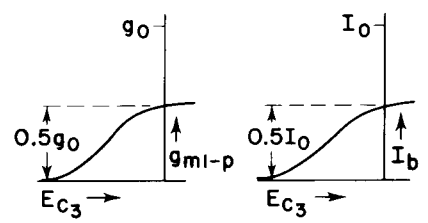
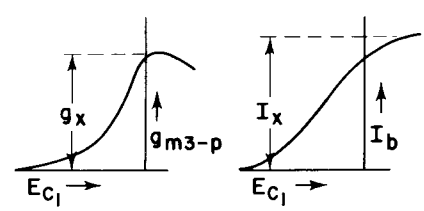
Herold's paper compared a triode, a pentode, an outer-grid-injection hexode, and an inner-grid-injection hexode. All four tubes were assumed to use the same geometry for the cathode and grid No. 1 structures in the tube. Fig. 13, from Herold's paper, shows the type of characteristics to be expected in each type, the relative conversion transconductances, and the equivalent noise microvolts on the grid at a frequency of 60 megacycles and a 4-megacycle bandwidth. The values for conversion transconductance and noise for the inner-grid-injection hexode are not readily evaluated in terms of the front-end geometry and depend largely on the transconductance that can be realized between grid No. 3 and the plate. Experience indicates that the noise values for the practical inner-grid-injection mixer are not likely to be appreciably better than on the outer-grid-injection tube. Both of them are too high to meet the requirements of modern television receivers.

Frequency-modulation broadcast receivers (88-108 megacycle band) have been built using pentagrid converters. Satisfactory results were obtained but fringe-area reception was considerably worse than would be obtained with a triode mixer.

The noise in a second-stage tube in a receiver is not important if the gain in the first stage is large enough. The contribution of the second stage to the noise factor of a receiver is shown by the following formula:

$$F_{1+2} = F_1 + (F_2-1)/G$$

where F_{1+2} is the noise factor of the two stages combined, F_1 is the noise factor of the first stage, F_2 the noise factor of the second stage, and G_1 is the available power gain of the first stage. The factors should be stated as actual ratios, not in decibels. If the second

TYPE OF FIRST DETECTOR	TYPICAL CHARACTERISTICS	APPROX. CONVERSION TRANSCOND.	APPROX. EQUIV. NOISE RESIST.	EQUIVALENT NOISE MICROVOLTS ON GRID*
SIG. +OSC		$0.28g_o$	$\frac{13}{g_o}$	8
SIG +OSC.		$0.23g_o$	$\frac{15+21\frac{I_o}{g_o}}{g_o}$	16
SIG. OSC.		$0.14g_o$	$\frac{120\frac{I_o}{g_o}}{g_o}$	32
SIG. OSC.		$0.28g_x$	$\frac{57\frac{I_x}{g_x}}{g_x}$	—

*ASSUMING $g_o = 15 \times 10^{-3}$ MHOS
$I_o = 30 \times 10^{-3}$ AMPS
$\Delta f = 4$ Mc

Figure 13. Comparison of Four Fictitious Mixers Assumed to Have Similar Cathode and Grid No. 1 Structures

stage has appreciable gain, noise from subsequent stages may be neglected. F_{1+2} is then very nearly the over-all noise factor of the receiver.

A typical television mixer pentode such as the 6CG8 will have noise factor of 14 db, i.e., a ratio of about 25, at 200 megacycles. If the rf amplifier has a noise factor of 6 db (a ration of 3.98) and a power gain of 100, then

$$F_{1+2} = 3.98 + 24/100$$

which is equivalent to 6.25 db. An rf amplifier stage with lower noise factor and higher gain than the values chosen may readily be obtained with some of the rf twin triodes now available. With such tubes, there appears to be little to gain in over-all noise factor by reducing the noise in the mixer stage of a vhf television receiver.

DESIGN OF MIXER TUBES

Triode Mixers

Triodes are not ordinarily used as mixers except in the frequency ranges from 20 to 1000 megacycles; they have largely been replaced by pentodes for frequencies below 200 megacycles. The design objectives for a triode mixer are very much the same as those encountered in the design of a good rf amplifier triode for the same frequency band. They include:

1. High transconductance
2. Low noise factor
3. Low capacitances
4. Low lead inductances
5. High resonant frequency at the tube input leads.

Ordinarily, if the triode mixer follows a high-gain rf amplifier, the mixer need not be exceptionally good for transconductance or noise factor; therefore, some compromises with cost can be made.

A good mixer tube would ordinarily have a medium value for mu, about 50-70, and a sharp cutoff.

High transconductance and low noise factor may be obtained by using small grid-to-cathode and grid-to-anode spacings and fine wire on the grids[14].

The capacitances are kept low by using a small cathode area with high current density, and by avoiding, so far as possible, incidental capacitances outside of the active electrode areas of the tube.

The lead inductances may be kept low in miniature tubes by keeping the mount height as low as practical, keeping the cathode lead short, and by using multiple leads to the electrodes[15].

At frequencies above four or five hundred megacycles, the use of disc-seal tubes is indicated for best performance, although miniature tubes have been used at frequencies up to nine hundred megacycles.

Pentode Mixers

The design of a pentode mixer has about the same requirements as that of the triode. The feedback capacitance from grid No. 1 to plate would, of course, be considerably lower than in the triode, but a pentode intended for mixer service only would not require as low a value for this capacitance as an rf pentode would for the same frequency range, because the gain would be lower and the output frequency would be different from the input frequency. In some cases, mixer tetrodes have been designed, but the operating conditions to obtain high plate resistance (and gain) in a tube without a suppressor grid are more restricted.

A number of combination tubes have been developed in recent years in which a triode oscillator and the pentode (or tetrode) mixer are combined in one nine-pin miniature bulb. In the design of such a tube, a careful study of the probable circuit layout is highly desirable to establish the best basing arrangement for short connections to the circuit[16]. Types have been made commercially available with and without shielding between the two units; but, for best results in most circuits, shielding adequate to prevent undesired coupling seems the better choice.

Outer-Grid-Injection Mixers

The design of the front-end (the cathode, grid, and first screen-grid region) of an outer-grid-injection mixer is very similar to that of an rf pentode. Because the oscillator drive is on grid No. 3, grid No. 1 can readily be made of the variable-mu type for remote cutoff without the need for excess oscillator drive voltage. The use of remote cutoff for automatic gain control will increase the noise factor of such a design over that of a sharp-cutoff tube but, at low frequencies, the noise factor can ordinarily be neglected.

A discussion of the electronic design problems of the region extending from grid No. 2 to grid No. 4 is given later.

The design of the region including grids No. 2, No. 3, and No. 4 is largely a choice of high screen current or high oscillator drive. If grid No. 3 is chosen to have a low amplification factor, the oscillator voltage required to swing this grid to cutoff will be high; if the amplification factor of grid No. 3 is made high, too much electron current will be turned back to grid No. 2 even at slightly positive bias voltages on grid No. 3, (that is, at the peak, positive oscillator swing) with the result that the maximum grid-plate transconductance and the conversion transconductance will be reduced.

To prevent the return through the screen grid of electrons turned back in front of grid No. 3, it is advisable to put a pair of shielding plates on grid No. 2, as shown in Fig. 11 for the 6SA7, or a variation of this structure[17]. This shield should be designed to let as much as practical of the initial beam of current from the region of the cathode pass through the slot into the region between grids No. 2 and No. 3 without any great portion of the electrons retracing their paths. The use of a grid No. 3 siderod directly opposite the slot tends to spread the beam and to reduce the number of electrons returning directly into the slot. The siderod also reduces microphonics by keeping the portion of grid No. 3 that controls the electron stream from vibrating. Trouble from

microphonics has been experienced with cylindrical control grids with two siderods and large spans of fine wire because the resonant frequency of the wire span may be in the audio range and, therefore, the space current may suffer an undesirable modulation by microphonism.

When a negatively biased grid is between two positive screen grids, it is possible to have trouble from "grid-blocking" if the secondary-emission ratio of the control-grid surface is sufficiently greater than unity and if there is considerable resistance in the grid circuit. The negatively biased grid may be accidentally driven more than a few volts positive either by too strong an oscillator drive at some point in the tuning band or by a positive surge voltage during band-switching. With high secondary emission and a high value of grid-leak resistance (perhaps 50,000 ohms or more), the grid may reach a stable operating condition at a positive voltage only slightly less than the screen-grid voltage. In this condition, the tube ceases to operate properly but can be restored to its proper operating voltages by turning off the set for a few minutes. This trouble can be eliminated by selecting tube materials and processing which will keep the secondary-emission ratio of the control grid low.

Inner-Grid-Injection Converters

The inner-grid-injection converter depends for its oscillator performance largely on the grid-to-plate transconductance of the front end, and for its conversion transconductance and gain on the design of the region between grids No. 2 and No. 3.

For the design of an oscillator section, the front end may be considered as a medium-mu triode with high perveance; the electrode spacings and cathode area (and dissipation requirements of the electrodes) must be balanced against the capacitances and transconductance requirements for the particular oscillator service. For operation in a hartley circuit with swinging cathode, the tube must require an oscillator voltage swing which is not too high or trouble will be experienced with degeneration of transconductance or the cathode may swing more negative than the signal-grid voltage.[18]

The collector plates on the screen grid should be positioned to permit as much as possible of the space current to go through the screen grid, but to catch practically all of the returning electrons.

The design of the region between grids No. 2 and No. 3 is discussed later in general terms. Experience has shown, however, that it is desirable to have a reasonably wide space between grid No. 2 and grid No. 3 for best control and best value of conversion transconductance, but there seems to be no way at present to determine the optimum spacing for a given set of voltages without a considerable number of tests. The use of finer wire and higher turns-per-inch can improve the performance of the control grid considerably, but mechanical problems and rising costs usually limit the extent to which these solutions are practical.

A curve of conversion transconductance vs. bias voltage on grid No. 3 at a fixed oscillator voltage usually shows a drop-off in transconductance after a peak value is reached (see Fig. 6). The bias voltage for this peak may vary from tube to tube and also in the same tube as a function of oscillator excitation. In seeking to maximize the conversion transconductance of the tube at given grid bias voltage and other operating conditions, the designer should be careful not to end up with a design in which a considerable portion of the product will have the peak value of conversion transconductance under negative bias voltage conditions on the signal grid, and with a considerable drop in conversion transconductance at zero bias. This type of characteristic has caused poor set performance in several types of converter designs in the past, and has been a source of customer complaints.

It is advisable to study the performance of the curve of conversion transconductance vs. bias voltage in grid No. 3 for a number of sample tubes of a new design in the region near zero bias voltage and to make sure that the design is not optimized for one condition at the expense of general usefulness.

Application Testing

Thorough application engineering tests are especially necessary on mixer tubes and converter tubes because so many factors are involved in their successful operation, and because the circuit constants play such a large part in their proper functioning. Some of the difficulties that have not been discussed above, but which require application testing are hum, off-resonance microphonism, and excessive variation in performance over the expected range of oscillator excitation.

CURRENT CONTROL BY A CONTROL GRID FOLLOWING A POSITIVE SCREEN GRID

The problem of current control by an ordinarily negative control grid placed between two positive screen grids requires consideration. The electrons enter the region between the first screen grid (grid No. 2), and the control grid (grid No. 3) with a velocity approximately equivalent in electron-volts of velocity to the potential between the screen grid and the cathode. In passing through the first screen grid and the grid No. 1 before it, some of the electrons acquire transverse velocity components due to the nonuniformity of the field near the grid wires. There are also small velocity variations both toward the anode and transversely, caused by the thermal distribution of velocities at the cathode. When the electrons are slowed down in front of the negative grid No. 3, the space charge formed there depresses the potential in the field of that region, and eventually space charge will set an upper limit on the current per square centimeter that can flow on through the control grid (grid No. 3) under a given set of voltage conditions.

The geometry and the grid voltages of grids No. 2, No. 3 and No. 4, and the incoming current through grid No. 2 determine the control characteristics of grid No. 3. The problem is to determine the portion of the incoming electron current I_0 that will pass on through

grid No. 3 instead of returning toward grid No. 2. A knowledge of this current distribution as a function of grid No. 3 voltage would enable calculation of the grid No. 3-to-plate transconductance on the assumption that the ratio of grid No. 4 current to plate current is constant. The cutoff voltage and the maximum plate current and transconductance at zero-bias voltage (or a slightly positive voltage) could also be determined for use in outer-grid-injection mixer design.

Unfortunately, no accurate analytical solution for the current distribution in such structures has been derived to include space-charge effects, transverse velocity effects, and secondary-emission phenomena at the screen grids. There have been numerous attacks on this problem, but a considerable amount of cut-and-try engineering is still necessary to be sure that a design is reasonably close to the optimum.

Some aspects of this problem will be discussed in order to give a general picture of the mechanisms involved. The problem has generally been attacked in one of two ways, by neglecting transverse velocities and concentrating on space-charge effects[19, 20], or by neglecting space charge and treating the electron optical problem[21, 22]. Recently, there have been some limited attacks on the problem in which space charge and transverse velocities were both included[23, 24]. In general, solutions have been confined to idealized parallel-plane structures.

When transverse velocities are neglected, the problem reduces to that of a parallel-plane diode with uniform initial velocities[19,20]. Consider such a diode with the first electrode at positive potential V_g and the second at positive potential V_a, less than V_g, and with the electrodes spaced apart a distance a (in cm). It is assumed that current can flow only perpendicular to the two electrodes. Let the current I_0 from the first electrode increase from zero. At first, all of the current will flow to the second electrode, the plate, but as space charge increases, the potential will be reduced in the space and, eventually, a potential minimum will be formed, but all of the current will continue to flow to the plate. When a high enough value of I_0 is reached, the potential minimum will drop to zero volts and a virtual cathode will be formed. When this condition occurs, part of the current I_0 will return to the first electrode without crossing the potential minimum and the rest will continue on to the plate. Any additional increase in I_0 will shift the virtual cathode further back toward the first electrode, and a larger fraction of I_0 will return.

Fig. 14 shows the type of characteristic that would occur in such an idealized case for two values of V_a/V_g. As I_0 increases, all of it continues on to the plate as I_a until a certain level of space current (three units for $V_a/V_g = 0.2$) causes the potential minimum to change to a virtual cathode and I_a drops to about 10 per cent (at point A). Under small variations of I_0 about point A, the virtual cathode still remains. The plate current I_a is much less than I_0 and decreases with increasing I_0. Once I_0 drops below point B, the virtual cathode vanishes and the hysteresis loop may be repeated. It will be seen that in this idealized situation, there is

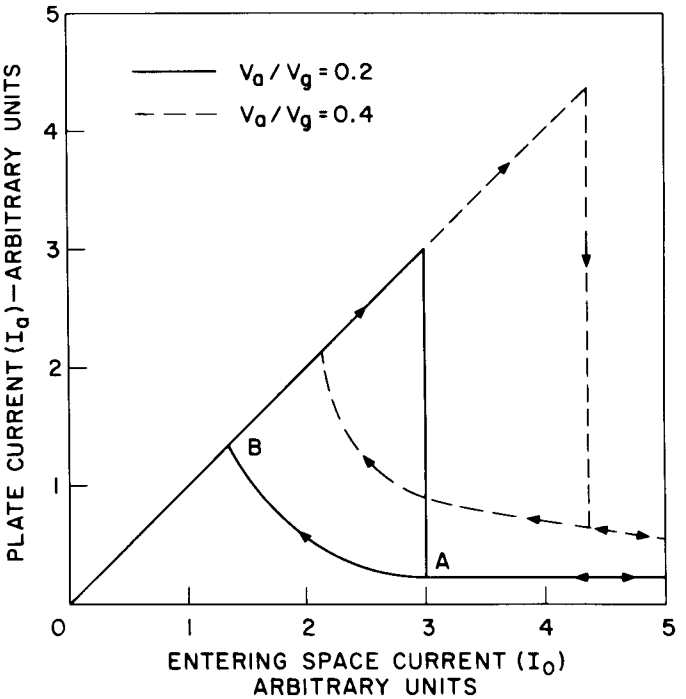


Figure 14. Curve of Plate Current vs. Space Current in an Idealized Parallel-Plane Diode which the Electrons Enter with a Constant Finite Velocity

no change in plate current with change in plate voltage until a virtual cathode forms during the plate-voltage excursion.

If the plate of the diode is replaced by a control-grid plane with a positive plate or screen grid beyond it, there will be no transconductance from control-grid variations until a virtual cathode formed.

Now, consider the electron optical problem in which space charge is neglected. Jonkers[21,22] has studied current distribution in multigrid tubes assuming (1) no space charge, (2) equal fields on both sides of the negative control grid, and (3) small ratios of grid-wire size to pitch on the control grid. His results indicate the influence of the degree of dispersion of the electron paths in passing through nonuniform fields in the planes of the grids; but his original assumptions are far from valid in many practical cases, and therefore, calculations based on his formulas may not be very useful.

In the electron-optical case, the distribution of currents is not influenced by space charge and, therefore, the geometry of the structure and the ratios of the various voltages are the controlling variables. In such a case, a curve of I_a vs. I_0 for various values of E_{c3}, would have the general shape shown in Fig. 15, in which the ratio of I_a/I_0 for a given grid bias voltage is independent of I_0.

Fig. 16 shows the type of I_a vs. I_0 curve that is obtained in an actual mixer tube, and represents the general shape encountered in such types. The values of I_0, the actual current flowing outward through the screen-grid (grid No. 2) had to be estimated in this curve; and so the general shape is significant but the ratios of I_a/I_0

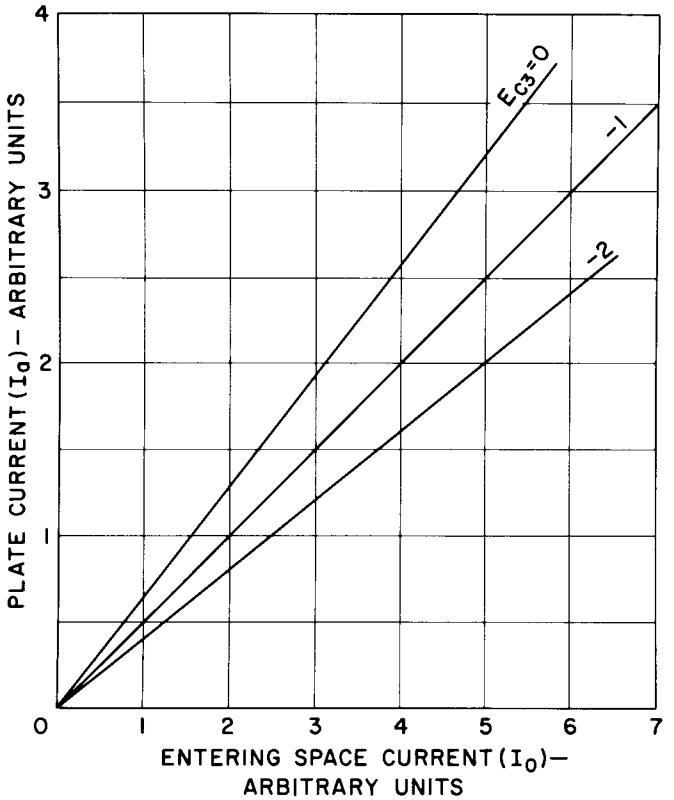


Figure 15. Idealized Curve of the Plate Current vs. the Entering Space Current (Through the Screen-Grid, Grid No. 2) in a Multigrid Tube in which the Grid No. 3 is Negative: (it is assumed that space-charge in the region between the grid No. 2 and grid No. 3 can be neglected and that the ratio of I_a/I_0 is changed only as the electron paths vary with changes in E_{g3})

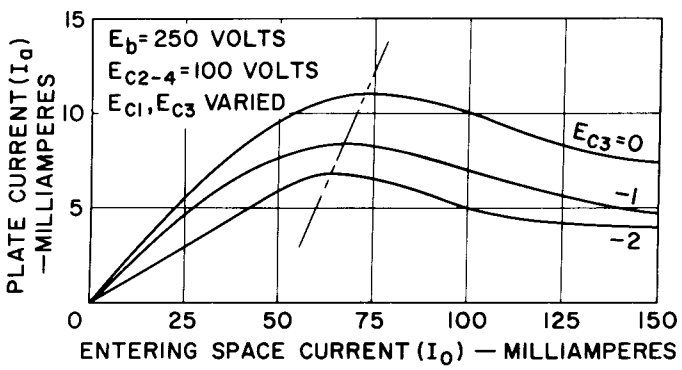


Figure 16. Curve of Plate Current vs. Cathode Current on the 6BE6, with Several Values of Grid No. 3 Bias

are not. It may be seen that the curves for different grid bias voltages fan out similar to those in Fig. 15, reach a maximum value, and fall off somewhat similar to those shown in Fig. 15, but without the sharp break or the hysteresis loop. The maximum value of I_a tends to shift toward the right with more negative control-grid bias voltage (that is, with a lower effective anode potential in the plane of grid No. 3), in correspondence with the idealized version of Fig. 14. This tendency of the peak I_a to shift is much more pronounced in parallel-plane structures such as the 6K8 than it is in the more-or-less cylindrical structure of the 6BE6.

In the design of a practical tube, the interest is usually in obtaining a high value of grid No. 3-to-plate transconductance per unit space current, or in obtaining a desired value of cutoff. General experience has been that the highest value of transconductance is reached somewhat sooner than the equivalent of point B in Fig. 14, that is, before a virtual cathode can possibly be formed. The use of fine wire grids, closely spaced to obtain relatively uniform fields, will improve the transconductance, because the effects of transverse velocities are reduced.

The maximum current density at which no virtual cathode can exist may be calculated from the following formula:

$$I_{am} = 2.33 \times 10^{-6} \ (E_{g2}^{3/4} + E_{ae}^{3/4})^2 \times \frac{1}{l_1^2}$$

in which I_{am} is in amperes per square centimeter, l_1 is the distance from screen grid (grid No. 2) to plate [or the equivalent anode in the plane of the control grid (grid No. 3)], E_{g2} is the voltage on the screen grid, and E_{ae} is the effective anode voltage in the plane of the control grid. Voltages are in volts; distances in centimeters.

The value of E_{ae} in the case under consideration may be calculated by the following formula:

$$E_{ae} = \frac{(E_2/\mu_{32}) + E_{g3} + (E_4/\mu_{34})}{1 + (1/\mu_{32}) + (1/\mu_{34})}$$

in which E_2 and E_4 are the effective potentials in the plane of grids No. 2 and No. 4, respectively, and may usually be taken to be equal to E_{g2} and E_{g4}, respectively, if the screening is adequate; $\mu_{32} = K_1 l_{23}$ (reverse mu); and $\mu_{34} = K_1 l_{34}$ (forward mu). K_1 is a function of the turns per unit length N and the diameter of the grid wire d used for grid No. 3. (See tables or chart in the article by L. G. Scholz in this book) This formula for E_{ae} may also be used to estimate the cutoff bias voltage of grid No. 3, although Jonkers[21] has found that a more detailed consideration of the field in the plane of the control grid will give more accurate results when space charge is not important.

It is interesting to note that considerable improvement in transconductance per unit plate current from a control grid following a positive accelerator electrode may be realized by taking special precautions in the design of the optics of the system to minimize the effects of transverse velocities. Such precautions were taken with design of the 6BN6 gated-beam tube,[25] intended primarily for limiter-discriminator operation.

REFERENCES

1. Langford-Smith, F. (ed.), Radiotron Designer's Handbook, 4th ed., bibliography pp. 1017-1019, RCA, Harrison (dist.) 1952
2. Radiotron Designer's Handbook, pp. 962-986 (see Ref. 1)

3. "IRE Standards on Electron Tubes: Definitions of Terms, 1957" Proc. IRE, vol. 45, page 990, July 1957
4. Radiotron Designer's Handbook, p. 965 (see Ref. 1)
5. Herold, E.W., and C.W. Mueller, "Beam-deflection Mixer Tubes for UHF." Electronics, vol. 22, pp. 76-80, May 1949
6. Radiotron Designer's Handbook, p. 967 (see Ref. 1)
7. "IRE Standards on Electron Tubes; Methods of Testing, 1950" Proc. IRE, vol. 38, pp. 943-945, August 1950
8. Dammers, B.G., J. Haantjes, J. Otte, and H. VanSuchtelen, Application of the Electronic Valve in Radio Receivers and Amplifiers, Book IV, pp. 329-335, Philips-Eindhoven (Netherlands), 1950
9. Radiotron Designer's Handbook, pp. 974-975 (see Ref. 1)
10. Herold, E.W., "An Analysis of the Signal-to-Noise Ratio of Ultra-High-Frequency Receivers," RCA Rev., vol. 6, pp. 302-331, Jan. 1942
11. Pan, W.Y., and D.J. Carlson, "Analytical Approaches to Local Oscillator Stabilization," RCA Rev., vol. 17, p. 534, Dec. 1956
12. Radiotron Designer's Handbook, p. 970 (see Ref. 1)
13. Herold, E.W., W.A. Harris, and T.J. Henry, "A New Converter Tube for All-Wave Receivers," RCA Rev., vol. 3, pp. 67-77, July 1938
14. Giacoletto, L.J., and H. Johnson, "UHF Triode Designs in Terms of Operating Parameters and Electrode Spacings," Proc. I.R.E., vol. 41, pp. 51-58, Jan. 1953
15. Harris, W.A., and R.N. Peterson, "Determination of Lead Wire Inductances in Miniature Tubes," RCA Rev., vol. 20, pp. 485-498, Sept. 1959
16. Smith, K.H., "Performance of Television Turret Tuners," Jour. of the Television Society, vol. 8, pp. 377-390, Jan.-March 1958
17. Radiotron Designer's Handbook, pp. 982-983 and Figs. 25.18 and 25.20 (see Ref. 1)
18. Radiotron Designer's Handbook, pp. 999-1000 (see Ref. 1)
19. Salzberg, B., and A.V. Haeff, "Effects of Space-Charge in the Grid Anode Region of Vacuum Tubes," RCA Rev., vol. 2, pp. 336-374, Jan 1938
20. Ivey, H.F., "Space Charge Limited Currents," Advances in Electronics and Electron Physics, pp. 170-184, Academic Press, New York, 1954
21. Jonker, J.L.H., "The Control of the Current Distribution in Electron Tubes," Philips Research Reports, vol. 1, pp. 331-338, Nov. 1946
22. Jonker, J.L.H., "The Internal Resistance of a Pentode," Philips Res. Rep., vol. 6, pp. 1-13, Feb. 1951
23. Walker, G.B., "Space-Charge Effects Between a Positive Grid and Anode of a Beam Tetrode," Wireless Engineer, vol. 22, pp. 157-169, 212-222, 276-281, April, May, June, 1945
24. Rodda, S., "Beam Tetrode Characteristics, the Effect of Electron Deflections," Wireless Engineer, vol. 23, pp. 140-145, May 1946.
25. Adler, R., "The 6BN6 Gated-Beam Tube," Proc. Nat. Electronic Conf., vol. 5, pp. 408-416, 1949

Design of AM and FM Detector Tubes

T. J. Henry

Harrison

DETECTORS

In radio, television, and similar communication systems, the transmitted signal is modulated by a modulation signal containing the information or intelligence to be transmitted. At the receiver, the operation of separating this modulation signal from the transmitted complex signal is called detection, or demodulation. By the time that this transmitted complex signal arrives at the detector in a receiver it has been somewhat degraded by: (1) distortion in the transmitter, (2) selective fading of sidebands during passage above the earth, (3) atmospheric noise and interfering signals from space, (4) noise in the receiver, (5) sideband clipping in the rf and if circuits, and (6) cross-modulation and modulation distortion from non-linearities in the rf, if, and mixer circuits. Most of these sources of degradation cannot be corrected for in the detector stage or after it, although standardized amplitude distortion in the transmitter and receiver can be compensated for to a large extent, and some degree of control of signal-to-noise ratio can be obtained in the detector.

The ideal demodulator would, in general, produce an output signal the same in content (subject to previous degradation) as the original modulation signal at the transmitter.

In practice, various distortions, such as harmonic distortion, amplitude distortion, and phase-delay distortion, occur in the demodulator. Economic considerations and the special requirements of any given system determine the design of the practical demodulator stage.

DETECTION OF AMPLITUDE-MODULATED SIGNALS

General Considerations

In amplitude modulation, the amplitude of the sinusoidal carrier-frequency wave is varied linearly in proportion to the modulating signal. See Fig. 1. The modulation factor m is defined in general for an unsymmetrical modulation as follows:

$$\text{For upward modulation } m = \frac{E_{max}-E}{E}$$

$$\text{For downward modulation } m = \frac{E-E_{min}}{E}$$

The maximum possible downward-modulation factor is

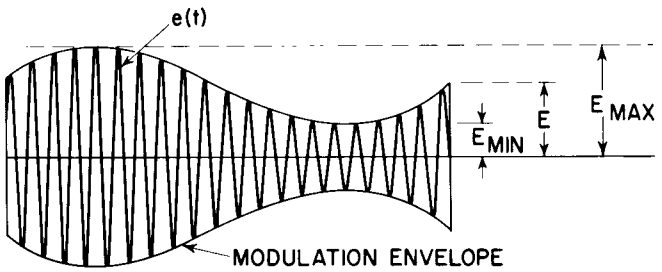


Figure 1. Amplitude-Modulated Carrier

1.0, but the upward factor is not limited. When the upward-modulation factor m exceeds 1.0, the signal is said to be over-modulated. Over-modulation is undesirable in amplitude-modulated (AM) radio because of the distortion that is introduced, but some over-modulation in a television signal can be tolerated during synchronization pulses.

In general, the amplitude-modulated signal has a carrier of frequency f_c modulated by a complex signal G(t). G(t) can be resolved by the Fourier method into a plurality of sinusoidal waves. The AM signal is physically equivalent to a carrier signal with symmetrical sidebands displaced from the carrier by plus and minus the frequency of each sinusoidal component of G(t). The pass-bands of the transmitter and receiver determine the maximum frequency component in the final information at the detector input.

The Diode Envelope Detector

The most widely used detector of AM signals is the diode rectifier. A typical radio-receiver detector circuit is shown in Fig. 2. This circuit is substantially the same as that used for a half-wave power-rectifier diode with capacitor-input filter, but the amplitude of the output signal varies with the modulation envelope, and steady-state conditions do not apply.

The diode detector circuit has been analyzed in the literature in many articles over the past thirty years; satisfactory design parameters of the circuits for various applications have been established.[1,2,3] Since most of the problems relate to the circuit rather than to the diode, they will not be covered in detail here.

The diode detector is widely used because it is an economical method of detection, not easily subject to

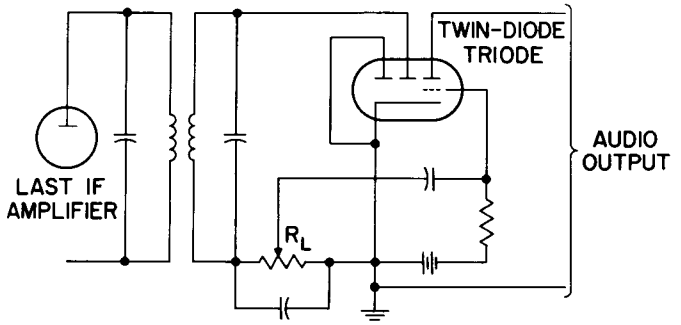


Figure 2. Diode Detector and Triode Amplifier for Amplitude-Modulation Radio Receiver

overloading; and, when properly designed, will give rectification that is linear enough for many purposes. It also provides a negative bias voltage that can conveniently be used for automatic gain control.

At low input levels, some distortion arises because the curve of plate voltage vs. plate current for the diode is not linear and, therefore, square-law output is appreciable at such levels. Above several volts input, this source of distortion is negligible for most purposes. Distortion at high modulation levels also arises because (1) the diode RC circuit cannot follow the envelope of the signal exactly at high modulation levels and (2) the ac load on a diode is, in general, lower in impedance to the modulation frequency than to direct current. As a result, a phenomenon known as peak-clipping takes place.

A typical television video-detector circuit is shown in Fig. 3. This circuit operates on the same principle as the radio detector, but it includes inductances added to control the shape of the video response curve, and uses a much lower value of load resistance, about 5000 ohms, than the radio detector. There are a variety of diode-detector circuits used for television receivers, [4, 5] but they are all much the same so far as the function of the diode tube is concerned.

In the design of the video detector, the television set designer is concerned with obtaining the proper video response, a reasonable rectification efficiency, low distortion, minimum radiation of spurious signals, and minimum cross-modulation between sound and picture. In color television, the chroma subcarrier adds to the problems of cross-modulation and distortion.

In most sets, the video detector serves as a frequency converter by beating the portion of the video if signal

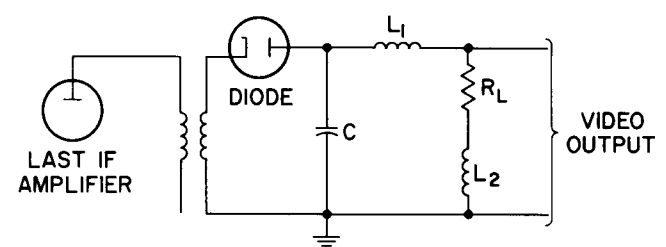


Figure 3. Series Diode Detector Stage for Video Detection in Television Receiver

containing the sound information against the picture if carrier signal to produce a 4.5-megacycle carrier frequency modulated with the sound information. [6] This intercarrier sound signal is then fed into the sound if stage and from then into the sound detector.

As will be discussed later, the vestigial-sideband signal used in television cannot be detected by an envelope detector without distortion. In practice, however, the defects of the diode type of envelope detector have not been serious enough to prevent its almost universal use.

Some special problems in the design of diode detector tubes will be covered later.

Other Envelope Detectors

There are a number of other types of envelope detectors, many using multi-electrode tubes and having the advantage of providing gain as well as detection. Two of these multielectrode detectors, the grid-circuit detector (commonly called the "grid-leak" detector) and the plate-circuit (or grid-bias) detector[1] were frequently used before 1930, but because of their limited range of output, and lack of a convenient source of automatic-gain-control voltage (agc) are little used today. The infinite-impedance detector[7] has possible special applications when low distortion at high modulation levels is important enough to justify the added expense of an extra amplifier tube for automatic gain control.

Another system of detection, exalted-carrier (or enhanced-carrier) reception, is also useful in reducing distortion for signals which are received at high modulation levels. [2] In this method, before detection of the signal, a local source of sinusoidal signal of carrier frequency is added to the received signal in proper phase. In effect, this addition produces a new rf signal of larger amplitude and of low modulation ratio that may then be detected in the diode rectifier stage with low distortion. Because of the complexity of providing a suitable local signal, applications of this circuit in radio are limited to specialized communication receivers.

Single-Sideband and Vestigial-Sideband Modulation

The amplitude-modulated carrier signal with symmetric sidebands requires a bandwidth of twice the maximum modulation frequency and a signal power equal to the constant carrier power plus the power in the sidebands. With complete sinusoidal modulation, the total power is one and one-half times the carrier power. All of the information to be transmitted is contained in either sideband and can be recovered at the receiver from just a single sideband by appropriate demodulating means. Elimination of the carrier and one sideband at the transmitter produces a considerable saving in power and in the frequency spectrum required.

In order to detect the single-sideband signal at the receiver, it is possible to: (1) introduce a local signal at the carrier frequency and then detect the mixture in an envelope detector, or (2) apply the single-sideband signal to a product detector, a circuit whose output contains the product of two input voltages, in this case, the single-sideband signal and a locally generated carrier-

frequency signal. Product detectors will be discussed in more detail later.

In many applications, such as normal monaural telephony, the local signal of carrier frequency does not need to be referred in phase to the original carrier. Where phase relations are important, some method of synchronization is necessary.

When a symmetrically modulated AM signal contains modulation components of low enough frequencies, complete elimination of the carrier and the unwanted sideband cannot be done in practical networks without excessive phase distortion. In such cases, vestigial-sideband transmission is frequently used. The standard monochrome (black-and-white) television channel (Fig. 4) uses this system with 4.5 megacycles in the upper sideband and about 1.0 megacycle of the lower sideband remaining.

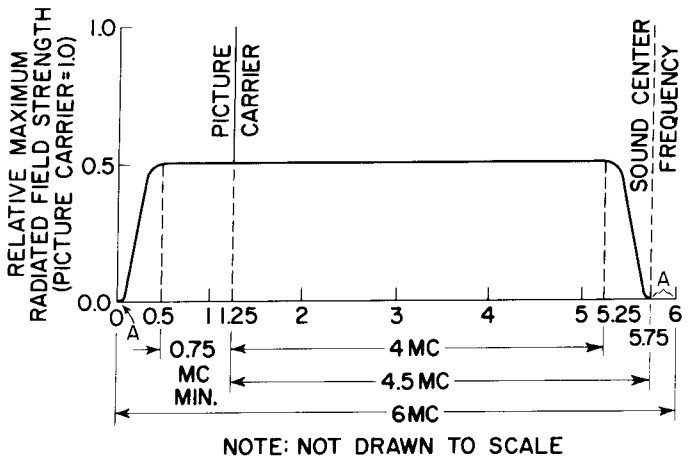


Figure 4. The Standard Monochrome Television Channel in the United States: the regions marked "A" must exhibit 20-db attenuation of the picture sidebands.

In color television, a similar but more complex vestigial-sideband signal is used (Fig. 5) with a slightly wider upper sideband and with a chrominance subcarrier at 3.579 megacycles. The chrominance signal contains two independent components of color information E'_g and E'_I that are present as independent chrominance-subcarrier modulations in quadrature.[8]

After an amplitude-modulated carrier signal with symmetric sidebands has been passed through a filter network to produce a vestigial-sideband signal, the envelope of the resultant asymmetric sideband signal is no longer linearly proportional to the original modulation. Detection by an envelope detector will result in a distorted signal.[9]

This distortion is a nonlinear function of the depth of modulation, and is accompanied by phase modulation of the carrier at the video modulation frequency. In monochrome television, these effects result in incorrect rendition of high frequency picture detail and an asymmetric transient response. In color television, in addition to defective rendition of brightness detail caused by the

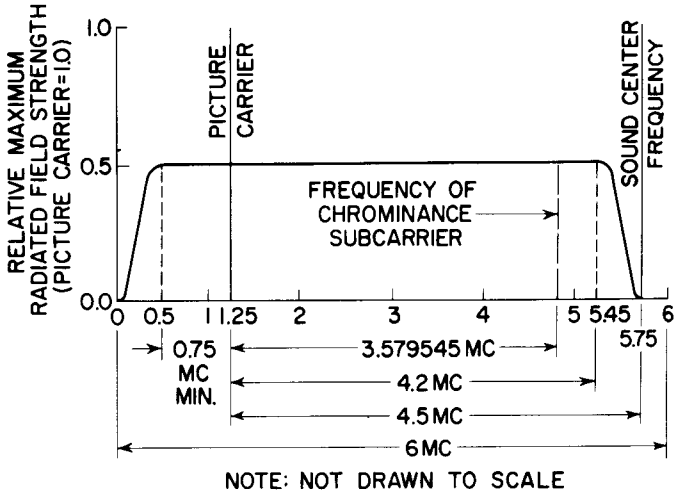


Figure 5. Color Television Channel

nonlinear envelope distortion, spurious color transients are produced by the phase modulation of the picture carrier.

The product-detector system can be used to demodulate vestigial-sideband signals without the distortion effects inherent in envelope detectors.

Product Detectors

The general principle of operation of the product detector will be illustrated here by a simplified mathematical approach. More extensive mathematical treatments may be found in the literature.[9, 10]

In the product detector, demodulation is obtained by multiplying together (1) an unmodulated carrier-frequency signal (usually locally generated) and (2) the modulated signal that is to be detected. Detection can be satisfactorily obtained whether the modulated signal has symmetrical sidebands, asymmetrical sidebands, suppressed carrier, or single sideband.

The ac output current of a tube with two input signal voltages can be expressed in the form of a Taylor's series in which the two input signal voltages are independent variables. If the output of such a nonlinear device is a function of two independent variables, there will generally be present in the output a second-order term (in the Taylor's series) which is proportional to the product of the two independent variables:

$$K \frac{\delta^2 i}{\delta e_2 \, \delta e_2} e_1 e_2$$

or $$K \frac{\delta g_{m1}}{\delta e_2} e_1 e_2 \qquad \text{when} \quad \frac{\delta i}{\delta e_1} = g_{m1}$$

Let $e_1 = E_1 \cos(\omega t + \Theta)$

$$e_2 = E_2 \cos \omega t + E_u \cos(\omega + \alpha)t + E_l \cos(\omega - \alpha)t$$

where ω is the angular carrier frequency, Θ is an arbitrary phase angle, and signal e_2 is modulated by a signal of angular frequency α. E_u is the voltage of the upper sideband component and E_l is the voltage of the lower sideband component. E_2, E_u, and E_l are independent, and e_2 represents the general case including as special cases signals with symmetrical sidebands, vestigial sideband, suppressed carrier, or single sideband.

If the cosine-wave values for e_1 and e_2 are put into the Taylor's series term shown above, and the usual trigonometric manipulations are performed, the output plate-current signal component resulting from this product term becomes:

$$i = K \frac{\delta g_{m1}}{\delta e_2}\Big[E_1E_2 \cos(2\omega t + \Theta) + E_1E_2 \cos\Theta + E_1E_u \cos(2\omega t + \alpha t + \Theta) + E_1E_u \cos(\Theta + \alpha t) + E_1E_l \cos(2\omega t - \alpha t + \Theta) + E_1E_l \cos(\Theta - \alpha t)\Big].$$

If terms of order $2\omega t$ can be filtered out, the remaining ac output component after further manipulation becomes:

$$i_d = K \frac{\delta g_{m1}}{\delta e_2}\Big[E_1(E_u + E_l) \cos\Theta \cos\alpha t + E_1(E_u - E_l) \sin\Theta \sin\alpha t\Big].$$

When e_1 is a locally applied carrier frequency of adjustable phase Θ, Θ may be chosen to be 0, $\frac{\pi}{2}$ or some out-of-phase value and for:

$$\Theta = 0, \qquad i_d = K\frac{\delta g_{m1}}{\delta e_2} E_1(E_u + E_l) \cos\alpha t.$$

$$\Theta = \frac{\pi}{2}, \qquad i_d = K \frac{\delta g_{m1}}{\delta e_2} E_1(E_u - E_l) \sin\alpha t.$$

Examination shows that the detected output signal i_d resulting from this product term in the Taylor's series is independent of the signal carrier amplitude E_2, whether one or both sidebands are present, and the output is a linear function of the sideband voltages E_u and E_l (that is, of the depth of modulation) even for off-axis detection.

In the general non-linear device, numerous higher-order terms will occur in the Taylor's expansion and these may contribute important distortion terms to the demodulated signal, provided their magnitude is large enough. [10]

A nonlinear device to serve as a product detector can be: (1) a diode or multielectrode tube in which both signals are applied to the same electrode, and in which $\delta g_{m1}/\delta e_2$ does not vanish; (2) a tube with two signal grids acting in series on a single electron stream; [11] or (3) a beam-deflection tube with the electron stream modulated in current amplitude by one control signal and in the direction and percentage of current to the output electrode by a deflection-control signal. [12] A combination of diodes or multielectrode tubes may also be used for product detection. [13] It should be noted that the two-input product device need not be nonlinear in the input-output characteristic from either input, because it is the dual action of two grids on the electron stream that produces the product term. Linear triodes or multielectrode tubes swept beyond cutoff will also provide nonlinear effects.

The product detector operates much the same in principle as the mixer tube in a superheterodyne frequency-converter stage. The major difference is that the oscillator frequency and the carrier frequency are the same, and consequently, the phase angle between these two input signals influences the output.

The method of analysis used by Herold for converters [14] may readily be applied to the product detector, provided the unmodulated signal e_1 is large with respect to the modulated signal e_2.

When a product detector is used with the nonmodulated signal synchronized by some means to the carrier of the received signal, it becomes a synchronous detector. Synchronous detection is required to separate the independent duplexed chrominance signals that are in quadrature on the color subcarrier of the standard American color-television signal. In this case, the synchronization is derived from the color-burst reference information in the transmitted signal. [15]

Synchronization is desirable in many other cases to eliminate the problems associated with phase and frequency drift that might otherwise arise. In some cases, such as in monochrome television, there may be some advantage to off-axis detection, even with synchronization, because of the improvement obtained in symmetry of transient response. [9]

Noise Advantage of Product Detectors

In the presence of thermal noise at the input to a product detector, the output will be a linear combination of (1) the in-phase components of the signals and (2) the noise present at the input; there will be no change in signal-to-noise ratio as a result of detection.

In the envelope detector at high signal levels, the output will be a linear sum of signal and noise, as in product detection; but when the signal grows weaker, the signal becomes suppressed at the output terminals and various cross-modulation and heterodyne effects arise between the noise and the signal.

Fig. 6 shows the signal-noise outputs of product and envelope detectors as functions of the signal-to-noise input ratios. In monochrome television, the measured advantage in product detection over envelope detection can be as much as 11 db for limit fringe signals.

There may be some disadvantage to product detection in television reception when impulse noise is present. There will be some whiter-than-white "sparkle" in the picture which can be eliminated by noise inversion or cancellation in the video circuits.

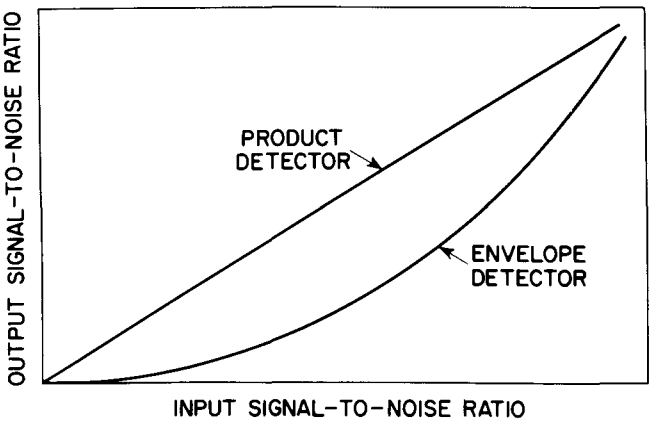


Figure 6. Comparison of Noise Response Characteristics of Product Detectors With Those of Envelope Detectors Showing the Poorer Signal-to-Noise Ratio in the Output With the Envelope Detector

The Design of Tubes for Amplitude-Modulation Detector Circuits

Diodes. The application requirements for diode detectors determine the characteristics required. For detection of conventional sound broadcasting with narrow bandwidths, there is no need to operate the diode with a load resistance lower than about 100,000 ohms, because lower values will give little improvement in quality and the problems of furnishing power to drive the detector will become greater. A diode that will supply one milliampere of dc current at 10 volts on the plate is adequate for this service. Such a diode does not require much cathode area or close spacing between cathode and anode. For economy of set design, diodes of this kind can readily be designed into a multipurpose tube having a triode or pentode and one or two diodes.

If a diode and a triode (or pentode) are in the same envelope (as in Fig. 2), and an ac signal is placed on the diode, some electrons will be accelerated towards the diode plate during the time that it is positive. Their number will depend upon the amplitude of the ac signal voltage. Some of these electrons, instead of being collected by the diode plate, may continue out into the space beyond the diode plate where they may be collected by the triode plate. The result is that a sort of triode is formed with the diode plate acting as the grid and the triode plate as the anode; the combination produces a distorted audio signal across the load in the triode plate circuit. This distortion is particularly objectionable in receivers in which the volume is controlled by a potentiometer arrangement across the diode load resistor R_L because the signal will not go to zero with the slider completely to the right; furthermore, the phase relations may be such that a null point occurs at some intermediate position of the slider.

This diode coupling effect may be made negligible by shielding the diode section of the tube with a grounded shield so that very few stray electrons can escape, or by arranging the tube geometry so that substantially no diode electrons can get to the triode anode. In a circuit similar to that in Fig. 2, except that the grid is grounded, the amount of diode coupling may be determined by applying an ac signal to the diode plate and reading the resultant ac voltage across the triode plate load. Any new design of multiunit tube intended for this service should be checked in this manner before the design is fixed.

Difficulties arising from coupling between the two diode units in amplitude-modulated radio applications in which each diode is used in a different circuit are apparently of little general importance.

Because the diodes are generally located on one end of the cathode, where temperature conditions are none too uniform, and changes in processing may have considerable effect, it is wise to make the design perveance at least twice that necessary to give the one milliampere at 10 volts mentioned earlier.

In addition, due to the position of the diodes and variations in tube processing, variations in contact potential occur which may be a source of trouble. In order to have any sensitivity at low signal voltages, it is necessary that the contact potential be such that the diode draw a small amount of current with no applied plate voltage. In some types of circuits, such as that of a diode-biased triode, it is important that the diode start to draw current before the triode grid. This type of circuit is used very little in present set design.

These diode-current starting points should be checked by a periodic quality test to control process changes which materially affect the diode-current cutoff.

In multiunit tubes, the temperature of the diode plates is so low that no trouble has been experienced with primary emission from the plates. Primary emission would produce a reverse current on the negative cycle of the applied ac voltage. In tubes where considerable cathode power is used, the diode plate temperature may become high enough to cause some primary emission and result in a reverse current. This reverse current need not be large to cause decreased sensitivity on small signals and should, therefore, be considered in the tube design. Process control or the use of plates with better radiating qualities, i.e., plates made of blackened, roughened material or with larger area, will prevent the reverse emission.

In television detectors, bandwidth considerations require that the diode load resistance be from 3000 to 7000 ohms; therefore, a high-perveance diode is needed to keep most of the drop across the load rather than in the diode. A tube with the characteristics of the 6AL5 twin diode appears to give adequate results. The 6AL5 will deliver a peak current of 54 milliamperes at 9 volts peak plate voltage. Before the introduction of the 6AL5, the 6H6 was used in this type of application, but it did not have a sufficiently high perveance to be completely satisfactory. The 6H6 will deliver about 12 milliamperes at 9 volts.

It is important to keep the cathode-to-plate capacitance low on a diode detector for television; to obtain low drop (or high perveance), it is, therefore, necessary

to use small cathode-to-plate spacings and small plate areas. The 6AL5 has cathode-to-plate capacitance of about 2.5 micromicrofarads for each diode.

In many uses of twin-diode tubes in television and frequency-modulation receivers, it will be found advantageous to shield the two diodes from each other to avoid electronic or capacitive coupling between the two units.

Two major causes of field trouble with diodes in television and frequency-modulation applications have been heater-cathode hum and variations in contact potential from tube to tube. Heater-cathode hum is due to heater-cathode leakage. If this leakage is too high, it may introduce undesirable amounts of hum in the output when circuits with the cathode off-ground are used. This problem is controlled largely by suitable tube processing and maintenance of the purity of insulating materials. The tube designer can also contribute by using a heater designed for relatively low-temperature operation. Contact-potential variations cause variations in sensitivity at low signal levels. Proper control of processes and materials, together with frequent spot checks on the product, appear to be the only remedies.

Triodes and Pentodes for Envelope Detection. There appears to be little general interest in the triode or pentode for envelope-detection application. No special design problems are apparent, and any sharp-cutoff tube satisfactory for operation at either the intermediate or radio frequency to be detected should serve the purpose. Triodes should probably have medium values of mu, i.e., between 20 and 70.

Product Detectors. As previously discussed, product detection may be obtained with (1) diodes, triodes, or pentodes, with the modulated signal and the beating signal on the same electrode, or (2) with tubes having two control electrodes affecting the same electron stream, such as pentodes with signals on grids No. 1 and No. 3, mixer hexodes or heptodes, or special beam-deflection tubes with dual control action. Triodes with one signal on the grid and one on the cathode are also used in color demodulators.

There appear to be no outstanding special requirements for diodes, triodes, or pentodes other than those required for envelope detection.

In tubes with two signal grids in series, it is desirable to have tubes that will handle reasonable signal levels without excessive distortion and that will not require large values of local carrier-signal voltage.

For use as the second detector in television receivers, it would be desirable to have a tube with a fairly high plate current (about 10 milliamperes), high values of transconductance on both signal grids, linear characteristics on the rf signal grid, low coupling between the two input circuits, and good operation at 45 megacycles. These specifications are desirable in order to provide gain, distortion, economy, and simplicity of circuitry competitive with those of other possible approaches.

The design of pentodes with input to both grids No. 1 and No. 3 is similar to the design discussed in the article on "The Design of Converter and Mixer Tubes" under Design of Mixer Tubes. The requirements for transconductance on the first grid and for cutoff on the third grid will largely determine the design. The allowable cathode power and interelectrode capacitances will set the spacings that must be used in the front end to obtain the required transconductance.

Mixer hexode and heptode designs will also follow the principles outlined in the discussion on Design of Mixer Tubes mentioned above, with the added problems of designing the additional array consisting of the second screen grid, the suppressor grid and the plate.

Beam-deflection tubes offer certain advantages for product-detector use over tubes designed to have two control grids in series. Less back coupling from the second control element to the first control element is encountered because the second control action on the plate current is obtained by switching the beam rather than by turning electrons back toward the cathode where they may influence the magnitude of the cathode current. (See the section on Coupling Effects in the article on "The Design of Converter and Mixer Tubes.") In addition, input loading is reduced because there are no returning electrons. Linearity of the deflection characteristics is good over most of the switching range; balanced circuits may be used for higher efficiency.

Two types of beam-deflection tubes suitable for operation as a product detector have been announced. The first design, the 6AR8 sheet-beam tube, was designed primarily for use as a color-demodulator tube, [12] but may be used for other product-detection schemes where a high value of deflector voltage is tolerable. Fig. 7 shows a schematic diagram of the cross-section of the 6AR8. The space current flows in two ribbon beams through a control grid (grid No. 1), a focus electrode, an accelerator grid, and between two deflecting plates (deflectors) towards a pair of anode or plate electrodes. The magnitude of the current in the beam is controlled by grid No. 1. The beam is directed to one anode or the other depending on the potential difference between the deflectors. A maximum dc cathode current of 30 milliamperes is permitted.

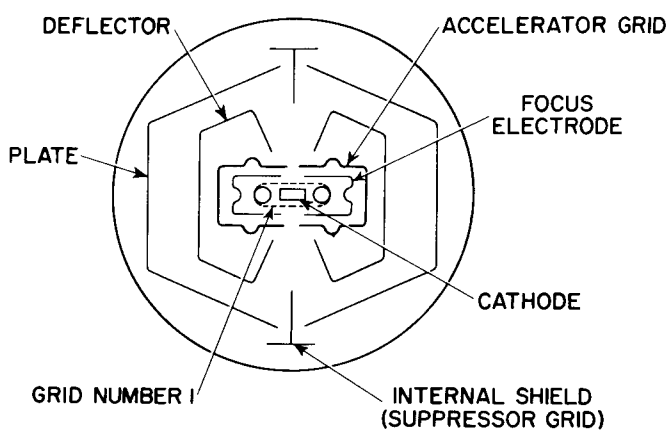


Figure 7. Schematic Cross-Section of the 6AR8

Fig. 8 shows the transfer characteristics of the 6AR8. The grid No. 1 to plate transconductance is 4000 micromhos at 10 milliamperes with plates and accelerator grid at 250 volts, and the deflectors at zero volts. Because it requires a change in deflection voltage of about 30 volts to shift 80 per cent of the beam current from one plate to the other plate, a large signal is required for efficient product detection. The deflector-voltage-vs.-plate-current curve is quite linear over the differential deflector-voltage range of plus 15 to minus 15 volts.

The design of beam-deflection amplifiers with ribbon beams has been discussed by Kilgore[16] and Pierce.[17] The transconductance from deflectors to one plate is determined by the length of path through the deflectors, the distance from a deflector to the intercepting edge, and the mean electron velocity in the deflection region. Limiting factors are allowable current density at the cathode, thermal velocities, and space-charge in the beam. The design of a conventional beam-deflection tube with much better deflection sensitivity than the 6AR8 within the confines of a miniature bulb and for a plate current of 10 milliamperes or more does not appear to be very practical. If the voltages are reduced to cut down the electron velocity and thus to increase the deflection sensitivity, thermal velocities and space-charge effects limit performance. The space within a miniature bulb does not afford room for improving deflection sensitivity by increased distances between deflectors and anodes.

Another beam-deflection tube recently developed, the RCA-7360, uses a different method of deflection and requires somewhat less voltage for switching the beam from one plate to the other.[18] It can be seen from the cross-sectional diagram shown in Fig. 9 that the unique feature of this type is the use of grid-shaped deflecting electrodes with the beam passing through the deflecting electrode to the collecting plates. A differential voltage between deflecting electrodes will switch the beam from one plate to the other. The characteristics of the 7360, as shown in Fig. 10, are quite similar in shape to those of the 6AR8. However, the transconductance from deflecting electrode to plate is increased, and less switching voltage is required.

According to conventional beam-deflection methods, the spacing between deflectors should be kept small to improve deflection sensitivity; but this requirement limits the excursion of the beam that can occur before the beam swings into a deflector, raises the deflection-circuit loading, and reduces the output current. In the 7360 design with porous deflecting electrodes, these electrodes are kept slightly positive at all times; but the current always passes through one or the other of

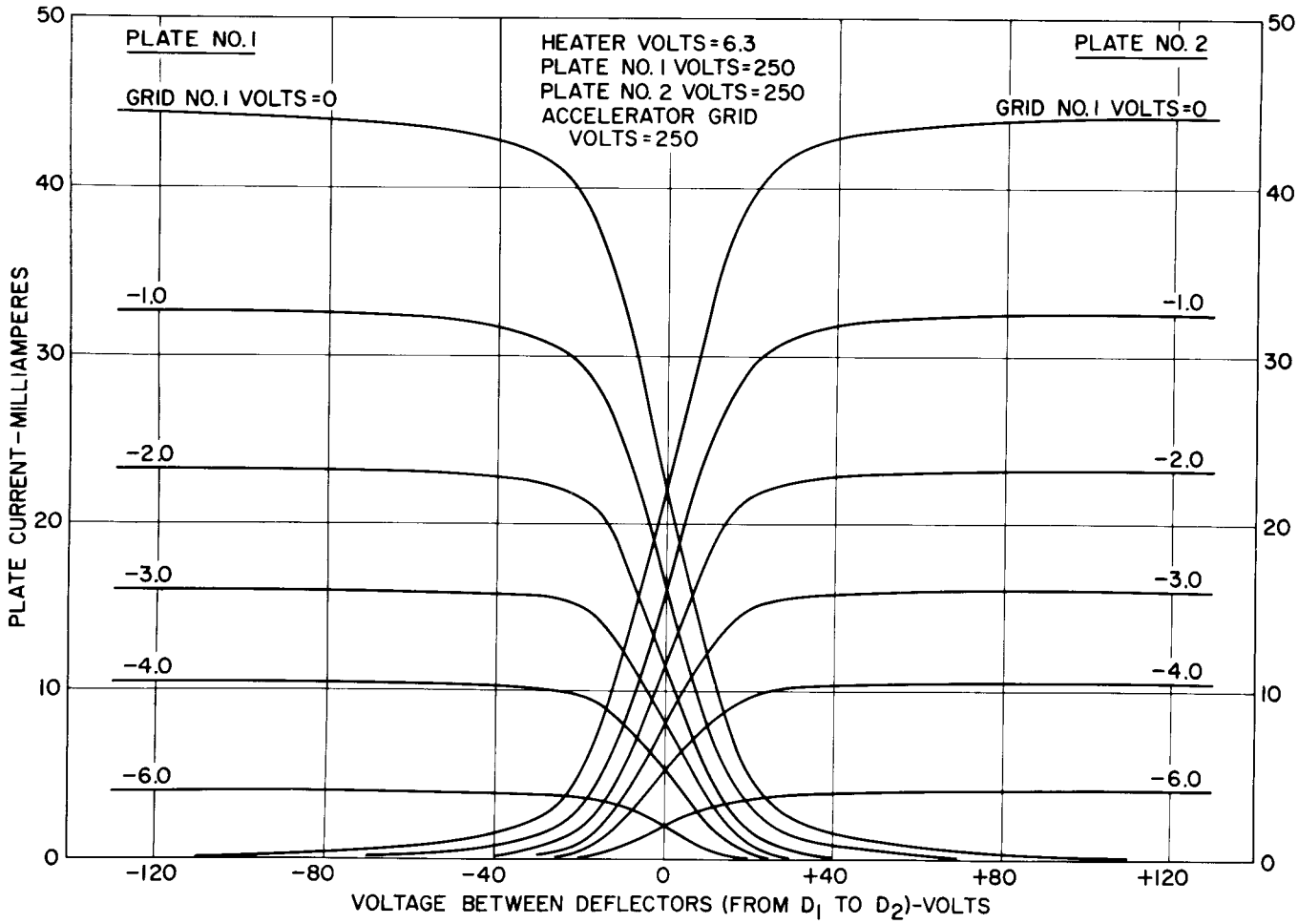


Figure 8. Average Transfer Characteristics of the 6AR8

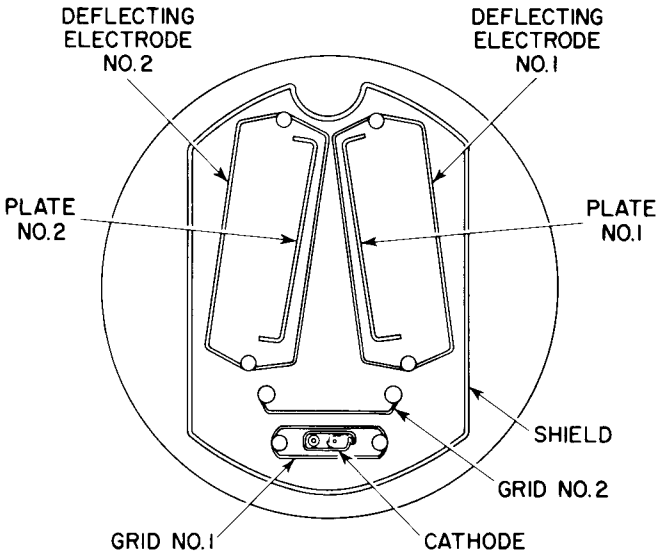


Figure 9. Cross Section of 7360 Beam-Deflection Tube

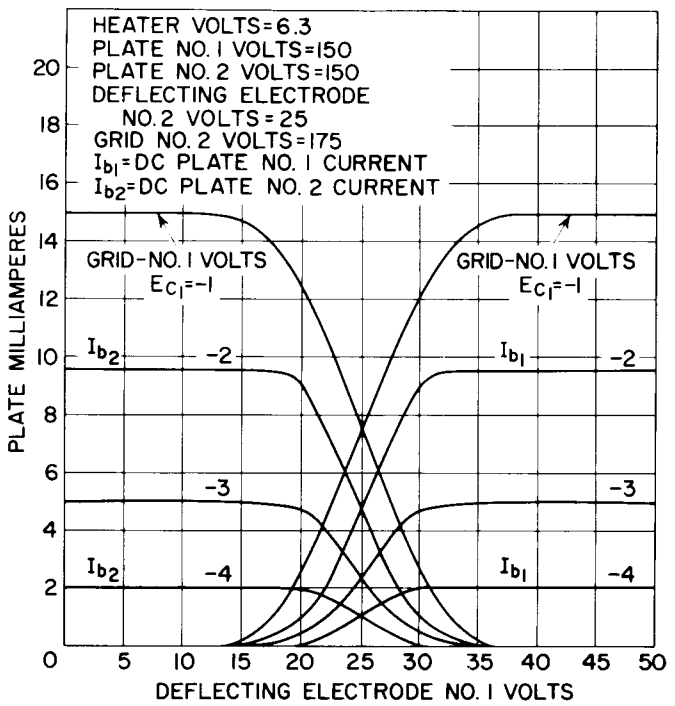

HEATER VOLTS=6.3
PLATE NO. 1 VOLTS=150
PLATE NO. 2 VOLTS=150
DEFLECTING ELECTRODE NO. 2 VOLTS=25
GRID NO. 2 VOLTS=175
I_{b1}=DC PLATE NO. 1 CURRENT
I_{b2}=DC PLATE NO. 2 CURRENT
GRID-NO. 1 VOLTS E_{C1}=-1
GRID-NO. 1 VOLTS E_{C1}=-1
PLATE MILLIAMPERES
DEFLECTING ELECTRODE NO. 1 VOLTS

Figure 10. Average Characteristics of Type 7360

them with only a small amount collected by the deflector even when the deflection signal is large. Consequently, smaller spacing between deflectors may be used for improved sensitivity.

DETECTION OF FREQUENCY-MODULATED SIGNALS

General Considerations

Frequency modulation has two useful characteristics compared to amplitude modulation. First, the power in the transmitted signal is independent of the degree of modulation; and second, atmospheric noise and other impulse noises such as ignition have less effect on the signal-to-noise ratio.

The instantaneous voltage e(t) produced by a carrier frequency modulated by a single sine wave is:

$$e(t) = E_c \sin \left(2\pi f_c t + \frac{\Delta f}{f_m} \sin 2\pi f_m t + \Theta_o\right)$$

where E_c = the maximum amplitude of the carrier voltage,
f_c = the carrier frequency,
Δf = the frequency deviation,
f_m = the modulation frequency,
Θ_o = a constant component of phase angle.

The ratio $\frac{\Delta f}{f_m}$ is known as the modulation index.

Transmission of more complex information may be resolved into a plurality of sinusoidal waves.

A mathematical analysis of the frequency spectrum of a frequency-modulated carrier shows that modulation by even a single sine wave results in an infinite number of sidebands separated by the modulation frequency f_m, but all sideband signals of significant amplitude are included in a band given approximately by[19]

$$\Delta F = 2(\Delta f + f_m)$$

In the case of a complex modulation, f_m should be the highest modulating frequency to be transmitted.

Frequency-modulated signals are transmitted with the amplitude held substantially constant; but selective fading and multipath interference in the air, nonuniform amplification over the passband in the receiver, atmospherics, and ignition noises introduce amplitude variations that the detector system must ignore for correct reproduction.

Detection of a frequency-modulated signal should, ideally, result in the reconstruction without distortion of the original modulating signal. The output amplitude of the device should vary linearly with the frequency deviation, and should be unaffected by accidental variations in the carrier-voltage level. There are several general systems of detection that give satisfactory performance over a limited range of carrier amplitudes. The set-designer's problem is to choose the most economical solution that meets the quality level requirements of the particular equipment.

Satisfactory operation using a frequency-modulation detector that is sensitive to amplitude modulation of the signal may be obtained by eliminating the amplitude modulation in a preceding "limiter" stage from which the output level is substantially constant for any input level above some threshold value.

A limiter stage is one in which the signal output varies more-or-less linearly with signal input until a threshold value is reached beyond which limiting action takes place and the output remains substantially constant for any input. Limiting may be obtained: (1) by using an if pentode stage in which the grid is biased by a grid-leak and capacitor with such circuit constants that the steady-stage gain varies inversely with the input; (2) by using a parallel-tuned LC circuit loaded by a diode negatively

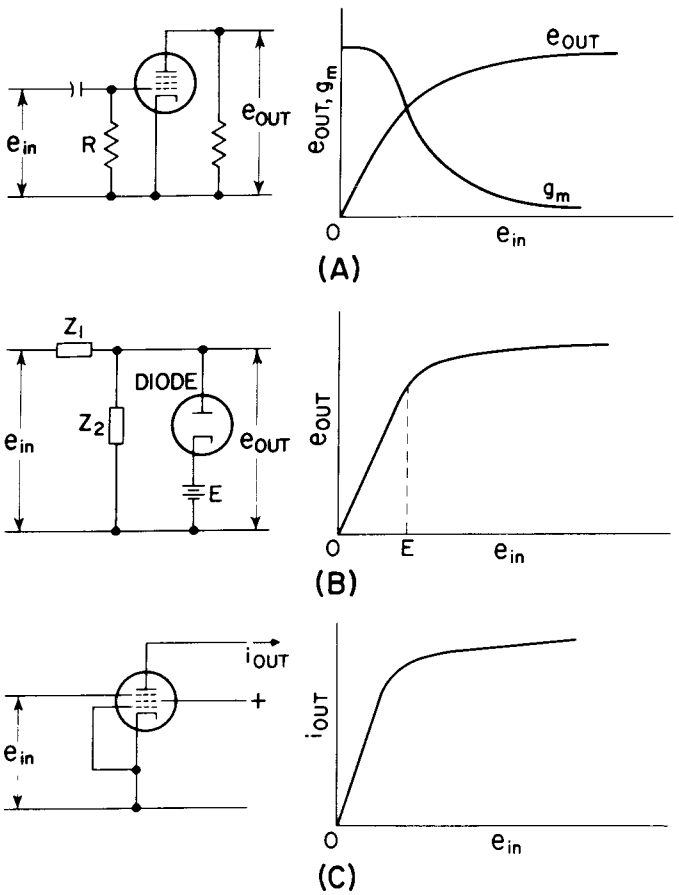


Figure 11. Three Methods of Amplitude Limiting: (A) Grid detection in a pentode in which the input signal on the RC circuit biases the tube more negatively with large signals and decreases the gain to maintain a substantially constant output voltage; (B) A diode negatively biased in shunt with a tuned circuit Z_2. When the signal input exceeds the bias voltage, diode current flows and loads Z_2. The difference in voltage between e_{in} and e_{out} is the drop across the impedance Z_1; (C) A multigrid tube with the input signal on an outer grid and the current limited by a preceding screen grid.

biased so that the voltage across the tuned circuit is limited by diode loading as soon as the signal is high enough for diode conduction; or (3) by using a multigrid tube in which an outer grid is used as the control grid, and the current is limited by a preceding screen grid. Fig. 11 shows these three types of limiters.

Practically all frequency-modulation detectors are somewhat sensitive to amplitude modulation. The range of amplitudes over which satisfactory operation may be obtained can be extended by using a final if stage that provides some limiting at high input levels. In a sense, the last if stage and the detector stage itself ought to be evaluated together as a detection system, even when a true limiter is not used.

Evaluation of this system would include consideration of the following points:

1. Threshold value of input level at some reference point (the grid of the last if stage, in general) below which detection does not occur or sensitivity to amplitude modulation is intolerably high.

2. Maximum input signal that can be satisfactorily detected without overload distortion becoming too high.

3. Degree of insensitivity to amplitude modulation over the useful range of signal levels, so-called AM rejection.

4. Gain of the system = $\frac{\text{output level of detected signal}}{\text{input level at the intermediate frequency}}$ at an established level of frequency deviation.

5. Distortion of the detected signal due to nonlinearity of the detector.

6. Sensitivity to impulse-noise sources such as static bursts and ignition noises.

7. The voltage output realized with a nominal value of frequency deviation. This consideration is important in determining if the power-output stage can be driven directly, or if a voltage-amplifier stage is required between the detector and output stages.

A useful method of evaluating the performance of frequency-modulation detector systems depends on the use of a frequency-modulation generator and an oscilloscope. The horizontal deflecting plates of the oscilloscope tube are connected to the modulating signal of the generator and the vertical deflecting plates are connected to the output voltage of the detector. An ideal detector would display the characteristic shown in Fig. 12A, independent of the carrier amplitude. A practical detector with perfect limiting would have a curve such as Fig. 12B with some nonlinearity and a tendency to flatten out at high deviation levels. If the signal is amplitude modulated as well as frequency modulated, and the AM rejection is imperfect, a display as shown in Fig. 12C may result. In this case, the AM rejection at center frequency is good, and a so-called "bow-tie" characteristic results. The degree of amplitude rejection may be measured from the width of the bow-tie. If AM rejection is equally poor over the frequency range, a broad band would be observed instead of the bow-tie.

As the amplitude of the signal is increased, eventually some sort of overloading will occur; a curve such as Fig. 12B or 12C will then start to distort and break up.

Types of Frequency-Modulation Detection Systems

A number of methods of frequency-modulation detection have been developed. Three basic methods are in general use:

1. The signal is applied to a network, the impedance of which depends upon frequency; the amplitude variations which result with frequency changes are rectified by one or more diodes.

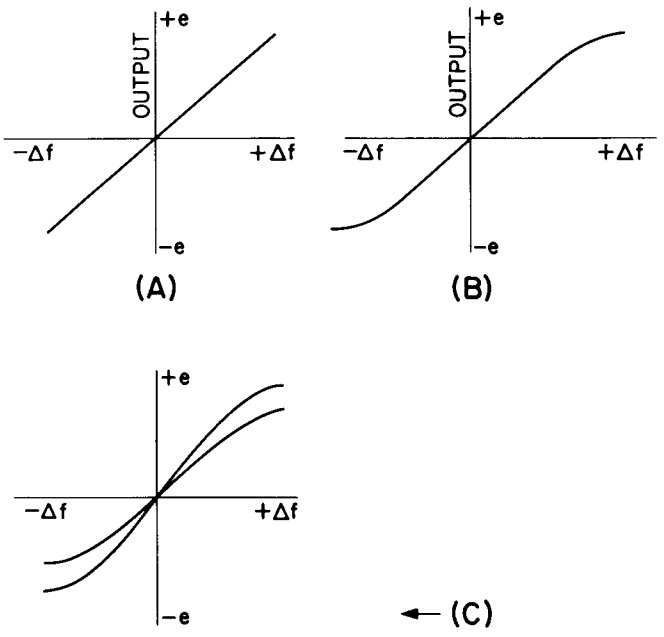


Figure 12. Oscilloscope Traces of Frequency-Modulation Detection Performance with Horizontal Sweep Synchronized with Modulation Signal: (A) ideal frequency-modulation detector; (B) frequency-modulation detector with nonlinearity; (C) frequency-modulation detector with nonlinearity and poor amplitude-modulation rejection (signal is both amplitude modulated and frequency modulated)

2. Two tuned circuits are coupled in such a way that the phase angle between them varies with the instantaneous value of the frequency. Detection is obtained by using a gating tube with two control grids acting on the same electron stream and, in this manner, obtaining plate-current pulses that are proportional to the phase difference. Because the phase-angle difference at center (carrier) frequency is 90 degrees, this method is often called quadrature-grid detection.

3. When an oscillator is oscillating weakly at its natural frequency, it can be pulled into synchronism with a strong external signal of nearly the same frequency. The dc current through the oscillator tube varies in a fairly linear manner with the frequency increment through which it is pulled; when a frequency-modulated signal does the pulling, detection is obtained by the "locked-oscillator" technique.

Detection by Amplitude Variations

The two most-used types of frequency-modulation detection by means of amplitude variations with frequency are the Seeley-Foster discriminator and the ratio detector. Fig. 13A shows the basic circuit of the Seeley-Foster discriminator and the response curve as a function of frequency. In this circuit, the ac voltage applied to each diode is equal at center frequency and the difference in amplitude across the two diodes increases almost linearly with frequency over a fairly wide range of frequency deviation. The ac signals are rectified in the two diodes and the instantaneous value of audio-frequency output (E_1 - E_2) is the difference of the two

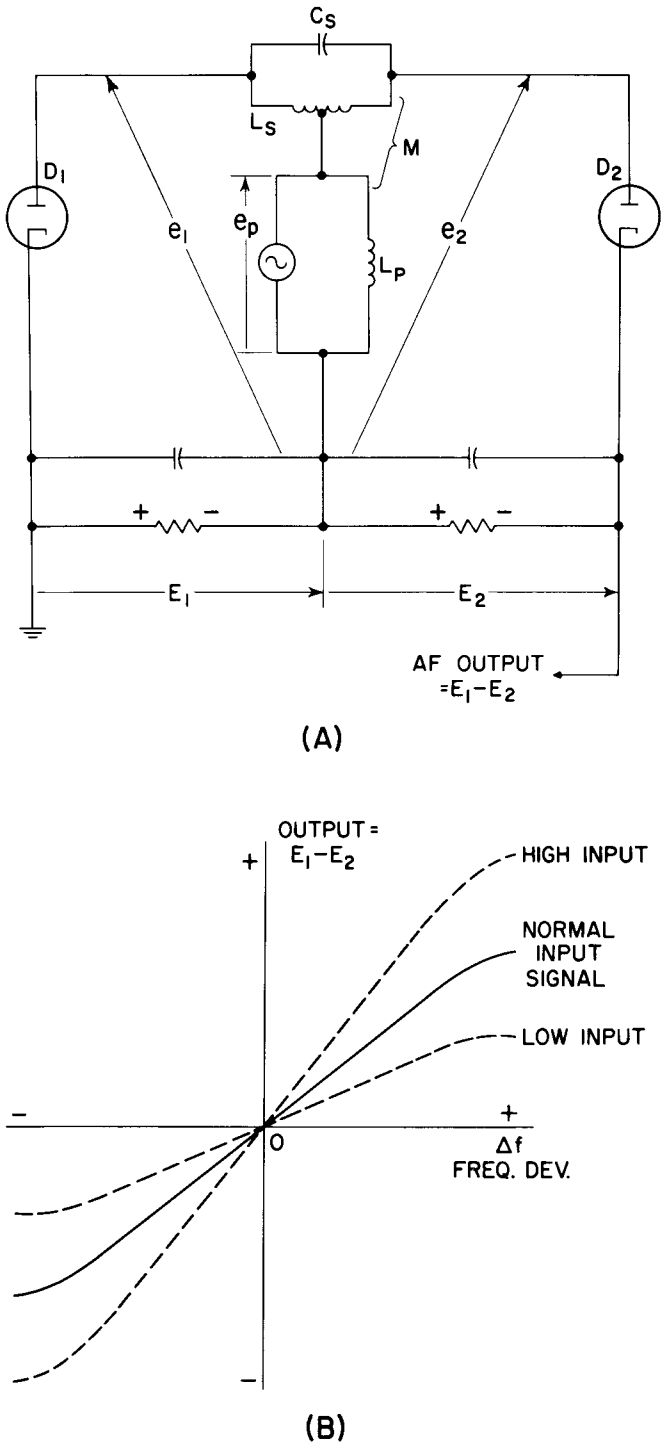


Figure 13. Basic Circuit of a Balanced Discriminator: the output characteristic in (B) shows the dependence of the output on the input signal amplitude

rectified e_1 and e_2 signals. As may be seen from Fig. 13B, this detector is sensitive to amplitude variations and requires a good preceding limiter stage for satisfactory operation.[20, 21]

The ratio-detector circuit shown in Fig. 14A uses a very similar network, but one of the diodes is reversed so that the rectified voltages across the two load re-

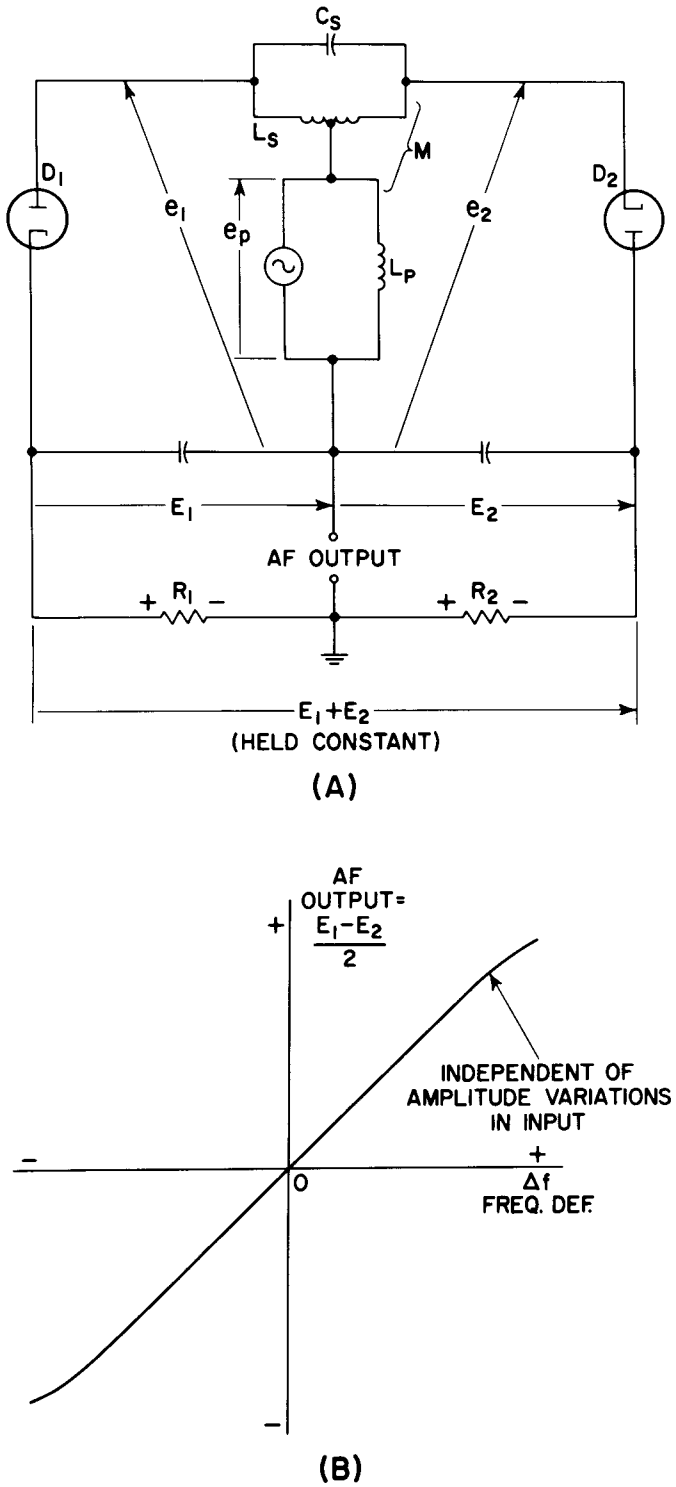


Figure 14. Basic Circuit of a Ratio Detector Using the Same Phase-Shift Input Circuit as in Fig. 13. The rectified voltage is stabilized so that the output can be independent of input amplitude

sistors add together rather than subtract,[22] with an output ($E_1 + E_2$). If ($E_1 + E_2$) is held substantially constant by a battery or a very large capacitor, the audio-frequency output taken off at the point shown will equal ($E_1 - E_2$)/2 and will be independent of amplitude so long as the ratio E_1/E_2 remains constant. (See Fig. 14B.)

The diodes required for satisfactory operation of these two detectors in a frequency-modulation receiver or the sound system of a television system should be on the order of the 6H6 or preferably the 6AL5 for perveance and capacitance. Balanced capacitances to ground (and to heater) and low hum are desirable features.

The ratio-detector gives excellent performance with acceptable distortion and amplitude rejection, provided it is carefully balanced. The ratio-detector transformer must have a well-balanced center tap, and is therefore usually made with a bifilar secondary winding.[22] This type of transformer is relatively expensive. The output voltage of the detector is low enough to require voltage amplification before the output stage.

Detection by Phase Variation with Frequency

When two inductance-capacitance (LC) circuits are tuned to the same frequency and coupled inductively or capacitively, a signal at resonance on one circuit will produce a signal in quadrature on the other circuit. If the signal frequency is varied, the phase of the voltages on the two circuits will shift as a function of frequency.[23] With a proper choice of Q values and couplings, the relationship may be made to be fairly linear up to a 30-degree phase shift from the quadrature position. If a tube has two control grids acting in series on the same electron stream and has a transfer characteristic for these two control grids as shown in Figs. 15A and 15B such that the plate current rises sharply with increasingly positive bias voltage, and if grid a and grid b are connected to each of two coupled circuits, the output current will be a pulse as shown in Fig. 16, because plate current will flow only when both grids swing positive. If the signal is frequency modulated, the phase difference between the circuits will change with frequency, and the pulse width will vary correspondingly. With a resistance-capacitance network in the plate of the tube, the pulses will be integrated and an audio-frequency voltage corresponding to the frequency modulation will appear in the detector output. If the plate current i_b shown in Fig. 15 were completely saturated and had a steep rise from zero, the detector would be insensitive to amplitude modulation of the signal.

Three different types of tubes have been designed for this type of detection, the 6BN6 gated-beam discriminator tube,[24] the 6BE7 "Ø" detector tube,[25] and the 6DT6 discriminator tube.[26]

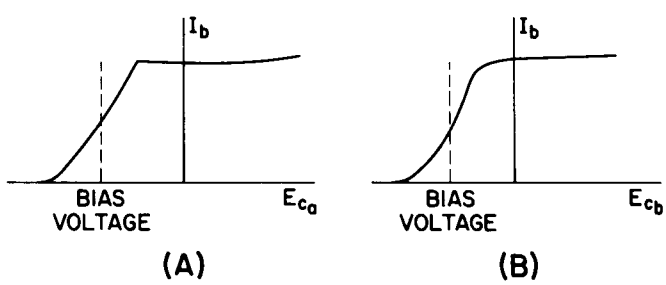


Figure 15. Transfer Characteristics for a Phase-Variation Detector Tube with Two Control Grids: (A) curve of plate current vs. voltage on grid a; (B) curve of plate current vs. voltage on grid b.

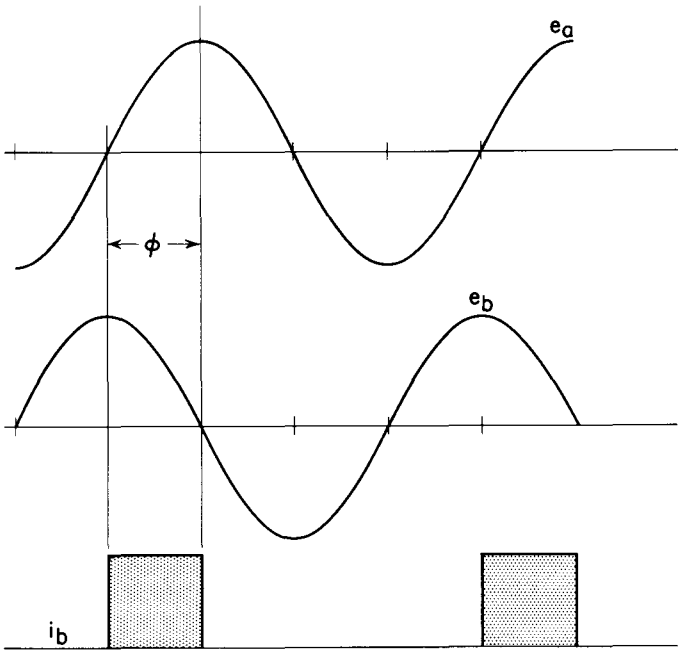


Figure 16. Principle of Operation of Phase-Variation Detection of Frequency-Modulated Signals. Plate current flows only when both e_a and e_b are positive. Phase angle between signals e_a and e_b is 90 degrees at center signal frequency, but changes as frequency is modulated.

The 6BN6 tube, shown in Fig. 17A, uses an electron beam focussed by electron-optics to reduce the effect of transverse velocities on the control characteristics of the second control grid and to achieve a steep rise of plate current (high transconductance). Both control grids are shielded from the cathode so that the space current saturates and produces a fairly flat-topped characteristic, as shown in Fig. 17B. Capacitive coupling between the two inputs through the tube is also very low. The circuit for the 6BN6 is shown in Fig. 17A. Coupling between the two tuned circuits is produced by "space-charge coupling" through the electron stream inducing a voltage on the second-control-grid circuit. The phenomenon of space-charge coupling is treated in the discussion of COUPLING EFFECTS in the article on "The Design of Converter and Mixer Tubes."

The "ϕ" detector tube, 6BE7, obtains saturation in the same manner as the 6BN6, but uses a conventional multigrid structure with seven grids. (See Fig. 18A.) Instead of a steep rise of the plate current, a symmetrical pulse is obtained by using variable-pitch winding on the two signal grids to produce the same shape of characteristic for both grids. The 6BE7 is designed to operate with inductive coupling between the tuned circuits and does not require the low interelectrode capacitances that are needed for the 6BN6 circuit. Because the 6BE7 has a characteristic with a rounded knee, as in Fig. 18B, it does not limit well until a high signal-input level is reached. The 6BN6 requires much less input than the 6BE7 for good limiting.

The 6DT6 discriminator tube has a comparatively simple construction and depends on an especially designed circuit for proper operation. The tube itself is

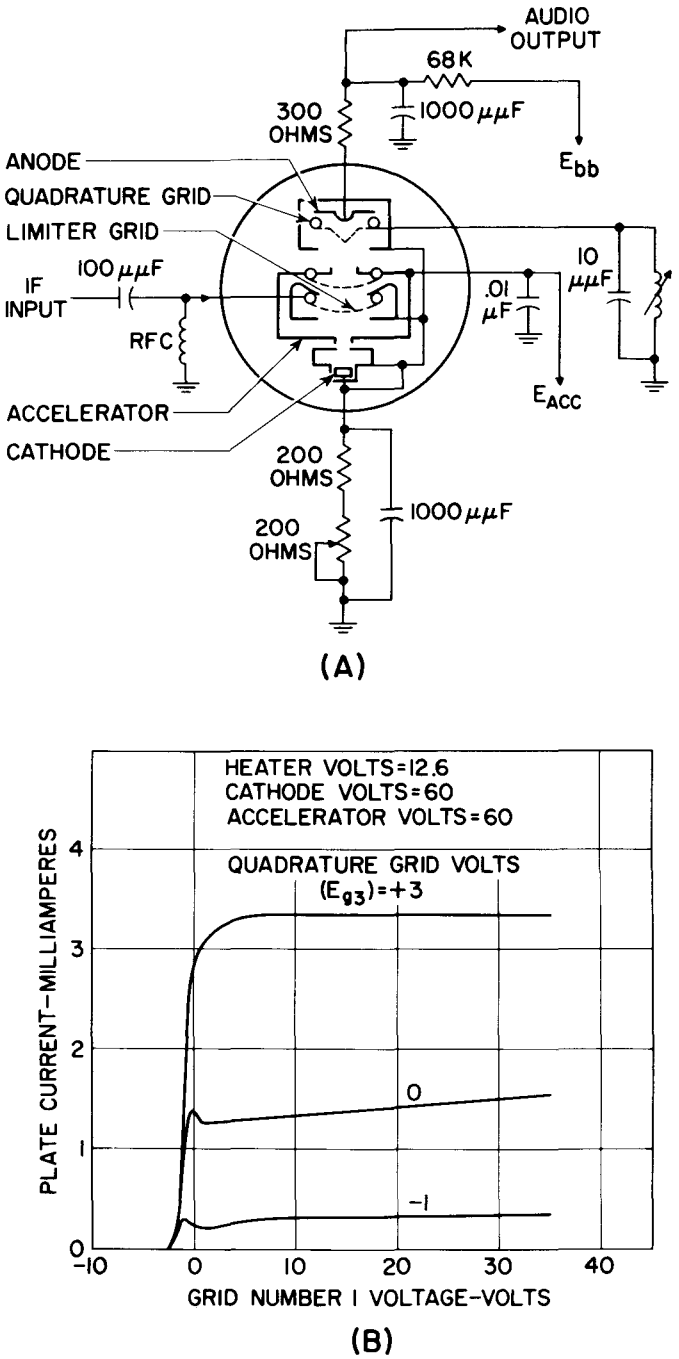


Figure 17. The Gated-Beam Detector Tube: (A) schematic diagram of typical frequency-modulation detector circuit and internal construction of the 6BN6; (B) average characteristic of 6BN6

a pentode with control on grids No. 1 and 3. The characteristics are shown in Fig. 19A and the circuit, in Fig. 19B. The 6DT6 in this circuit functions at low signal levels as a locked-oscillator detector and at higher levels as a gating-tube phase-shift detector. When no signal is applied to the circuit, the tube oscillates because of energy fed back through the capacitance between grids No. 1 and No. 3, due to the voltage induced by space-charge coupling of grid No. 1 to grid No. 3. The magnitude of this oscillation is limited, because the space-charge-coupling voltage tends to saturate when

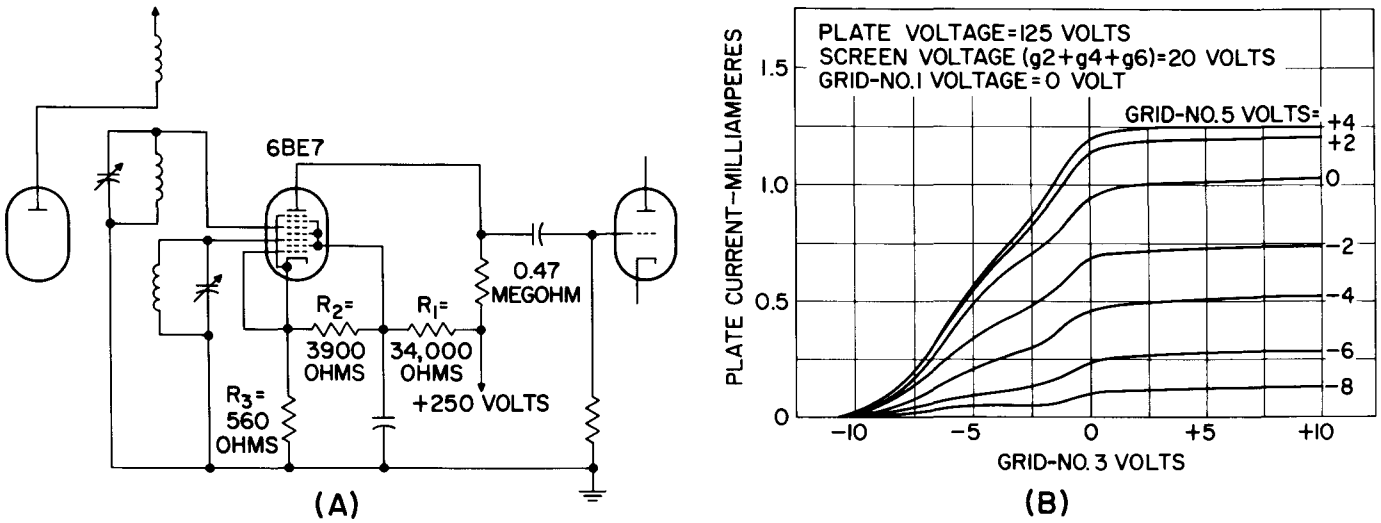
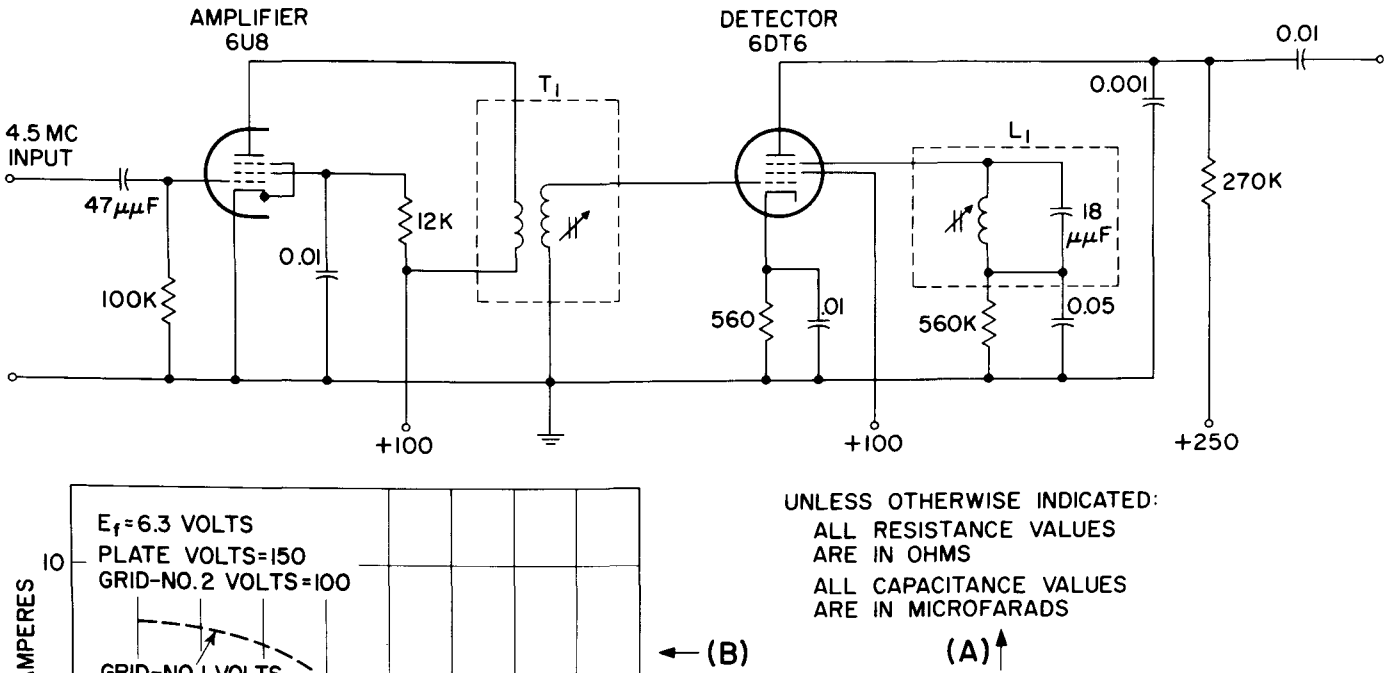


Figure 18. The ϕ-Detector Tube 6BE7: (A) a typical circuit for the 6BE7 as a frequency-modulation detector using inductive coupling between the two quadrature grids: (B) typical transfer characteristics of the 6BE7



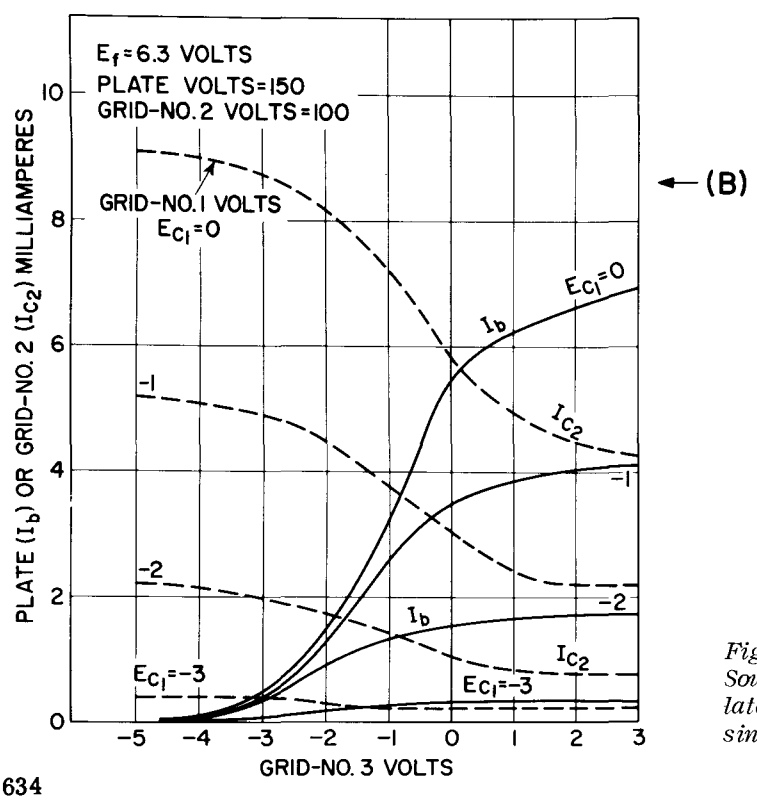


Figure 19. The Locked Oscillator Quadrature-Grid Sound Detector Tube 6DT6: (A) typical locked-oscillator and quadrature-grid detector circuit using a single-tuned input transformer; (B) average characteristics of the 6DT6

the space current through grid No. 2 becomes large enough to produce a virtual cathode between grids No. 2 and No. 3.

If a signal of increasing amplitude is applied to the input of the 6U8 driver tube, the signal on the input to the 6DT6 will reach a threshold level where it can lock the oscillator signal into synchronism and frequency-modulation detection will take place. With further increases in signal level, the oscillation ceases altogether and the phase-dependent width of the plate-current pulses will produce the detector action. The grid-No.-1-voltage-vs.-plate-current transfer characteristic of the 6DT6 does not show saturation; limiting in the grid No. 1 circuit arises from grid-current loading of the input circuit and from some limiting in the preceding stage.

The 6DT6 was designed primarily for use in television intercarrier sound systems at 4.5 megacycles with 25-kilocycle maximum deviation and would not necessarily work as well at the standard frequency-modulation intermediate frequency of 10.7 megacycles with 75 kilocycles maximum deviation.

The principles of design of tubes for this type of operation are discussed under the heading of DESIGN OF MIXER TUBES in the article on "The Design of Converter and Mixer Tubes."

REFERENCES

1. Langford-Smith, F., (ED.) Radiotron Designer's Handbook, 4th ed., pp. 1072-1082 and bibliography p. 1138, RCA, 1952

2. Landee, R. W., D. C. Davis, and A. P. Albrecht, Electronic Designers Handbook, McGraw Hill, New York, 1957

3. Wheeler, H. A., "Design Formulas for Diode Detection," Proc. I.R.E., Vol. 26, pp. 745-780, June, 1938

4. Deutsch, S., Theory and Design of Television Receivers, McGraw Hill, New York, 1951, chapter 6 pp. 191-199

5. Fink, D. G., Television Engineering, pp. 634-635 McGraw Hill, New York, 1952

6. (a) Fink, D. G., Television Engineering, pp. 426-430 (see Ref. 5)

   (b) Deutsch, S., Theory and Design of Television Receivers, ch. 14, pp. 456-473 (see Ref. 4)

7. Babcock, W. E., "Characteristics of AM Detectors," Audio Engineering, Vol. 35, No. 6, pp. 9-11 25-27, July 1951

8. Bailey, W. F., and C. J. Hirsch, "Quadrature Cross Talk in NTSC Color Television," Proc. I.R.E., Vol. 42, No. 1, pp. 84-90, Jan. 1954

9. Murakami, T., and R. W. Sonnenfeldt, "Transient Response of Detectors in Symmetric and Asymmetric Sideband Systems," RCA Rev., vol. 16, pp. 580-611, Dec. 1955

10. Livingston, D. C., "Theory of Synchronous Demodulation as Used in NTSC Color Television Receiver," Proc. of I.R.E., vol. 42, pp. 284-287, Jan. 1954

11. Pritchard, D. H., and R. N. Rhodes, "Color Television Signal Receiver Demodulators, "RCA Rev., vol. 14, No. 2, p. 205, June 1953

12. Adler, R. B., and C. Huer, "Color Decoder Simplifications Based on a Beam-Deflection Tube, "Trans. I.R.E., PGBTR-5, pp. 64-70

13. Murakami, T., and R. W. Sonnenfeldt, "Transient Response of Detectors in Symmetric and Asymmetric Sideband Systems," RCA Rev., vol. 16, pp. 580-611; Dec. 1955, Appendix III, p. 609.

14. Radiotron Designer's Handbook, pp. 964-966 (see Ref. 1)

15. Richman, D., "Color-Carrier Reference Phase Synchronization Accuracy in NTSC Color Television," Proc. I.R.E., vol. 42, pp. 106-133, Jan. 1954

16. Kilgore, G. R., "Beam-Deflection Control for Amplifier Tubes," RCA Review, vol. 8, pp. 480-505, Sept. 1947

17. Pierce, J. R., Theory and Design of Electron Beams, ch. 8 and 9, Van Nostrand, New York, 1954

18. Knight, M. B., "A New Miniature Beam-Deflection Tube," RCA Rev., vol. 21, pp. 266-289, June 1960

19. Landee, R. W., D. C. Davis, and A. P. Albrecht, Electronic Designers Handbook, page 5.28, McGraw Hill, New York, 1957

20. Foster, D. E., and S. W. Seeley, "Automatic Tuning," Proc. I.R.E., vol. 25, pp. 289-313, March 1937

21. Sturley, K. R., Radio-Receiver Design, Part 2, Chapters 13 and 15, Chapman and Hall, London, 1945 (Wiley and Sons, New York, 1945)

22. Seeley, S. W., and J. Avins, "The Ratio-Detector," RCA Rev., vol. 8, pp. 201-236, June 1947

23. Radiotron Designer's Handbook, page 422, (see Ref. 1)

24. Adler, R., "The 6BN6 Gated-Beam Tube, Part I;"

    Haase, A. P., "The 6BN6 Gated-Beam Tube, Part II;"

    Proc. of Nat. Electronics Conf., vol. 5, pp. 408-416, 416-426, Chicago, 1949

25. Jonker, J. L. H., and A. J. W. M. van Overbeek, "The O-Detector, a Detector Valve for Frequency-Modulator," Philips Tech. Rev., vol. 11, pp. 1-11, July 1949

26. Avins, J., and T. Brady, "A Locked-Oscillator Quadrature-Grid FM Sound Detector," RCA Rev., vol. 16, pp. 648-655, Dec. 1955

Vacuum-Tube Oscillators

W. E. Babcock

Harrison

OPERATION AND ANALYSIS OF TYPICAL OSCILLATORS

Oscillator Operation

In an amplifying device, such as a vacuum-tube amplifier, a steady, dc source of power is converted into periodically varying, or alternating power in the amplifier output. This process is normally controlled by a separate, external source of power converted in the grid circuit. However, most amplifiers can be arranged in such a way that the controlling, or exciting, power is obtained from the amplifier output. Such a device is known as an oscillator. Such feedback types of operation are not limited to devices containing electron tubes. For example, when a telephone receiver is held close to the mouthpiece of the transmitter, howling or singing results. In this case, the transmitter acts as an amplifier, while the receiver feeds part of the alternating power back through the air to the transmitter to cause sustained self-excited oscillations.

Sinusoidal Oscillators

From the dc supply, the vacuum-tube oscillator can produce alternating power with almost any desired stability of frequency and purity of waveform. It can generate power at frequencies from a few cycles per second up to several billion cycles per second and produce power outputs up to hundreds of kilowatts. The major application of sinusoidal oscillators to be considered here is for the generation of very small amounts of power for rf signal generators or for production of heterodyning signals for superheterodyne radios and television receivers. However, the basic principles to be discussed can also be applied to larger power applications such as transmitters, diathermy machines, etc. Another type of sinsoidal oscillator, the RC oscillator, is used in test equipment for producing audio-frequency power and will also be discussed in this section.

Before the various types of oscillator circuits are considered, some fundamental principles of oscillator operation will be examined. The simplest form of oscillator circuit consists of a parallel combination of inductance and capacitance as shown in Fig. 1. If capacitor C is given an electrical charge, the energy contained in the capacitor is equal to $1/2\ CE^2$, where E is the maximum potential across the capacitor plates. At the instant when E is at its maximum value, the circulating current in the circuit is zero. The capacitor then discharges through the inductance L. At the instant when the current through L is a maximum, the voltage across the capacitor is zero. The energy stored in the inductance at this time is equal to $1/2\ LI^2$, where I is the maximum value of current flow. The magnetic field then collapses, and energy is transferred back to the electric field of the capacitor. If there are no energy losses, either by radiation or by heat dissipation in circuit resistance, this process repeats itself indefinitely.

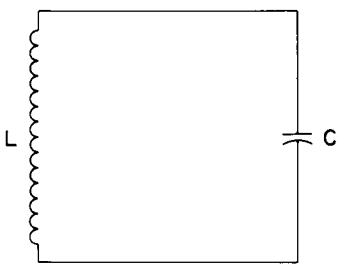


Figure 1. Simple LC Oscillator Circuit

Because the total energy stored in L and C must be the same, the expressions for stored energy given above may be equated:

$$1/2\ CE^2 = 1/2\ LI^2 \qquad (1)$$

The maximum potential, E may be expressed as:

$$E = \frac{I}{\omega C} \qquad (2)$$

where ω is 2π times the resonant frequency f of the oscillatory circuit.

Substitution of this quantity for E in Eq. (1), gives

$$\frac{1}{\omega^2 C} = L \qquad (3)$$

or $\omega^2 LC = 1$

And in terms of frequency

$$f = \frac{1}{2\pi\sqrt{LC}} \qquad (4)$$

where f is the natural resonant frequency of the oscillations occurring in the circuit.

In a practical circuit some losses always exist, both in the form of an actual circuit resistance loss and in the form of radiated energy. The equivalent circuit more closely representing an actual oscillator circuit is shown in Fig. 2.

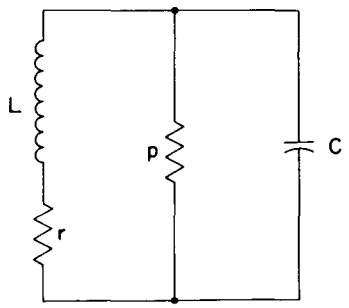


Figure 2. Equivalent Circuit for Figure 1

If the circuit in Fig. 2 is analyzed by Kirchoff's laws, the following equation is obtained for the current in any branch of the circuit:

$$\frac{d^2i}{dt^2} + \left(\frac{r}{L} + \frac{1}{pC}\right)\frac{di}{dt} + \left(\frac{r+p}{LpC}\right) i = 0 \qquad (5)$$

where r and p are resistances and L and C are the inductance and capacitance in the circuit, respectively.

A solution of Eq. (5) for current yields

$$i = Ae^{-\frac{1}{2}\left(\frac{r}{L}+\frac{1}{pC}\right)t} \sin \omega t \qquad (6)$$

where A is a constant.

The frequency of oscillation is given by

$$f = \frac{1}{2\pi}\sqrt{\left(\frac{r+p}{p}\right)\left(\frac{1}{LC}\right) - \frac{1}{4}\left(\frac{1}{pC}+\frac{r}{L}\right)^2} \qquad (7)$$

The Exponential Factor. Eq. (6) shows that the currents in the circuit may decrease, remain constant, or increase, depending on the sign or magnitude of the quantity $(r/L + 1/pC)$. If $(r/L + 1/pC)$ is positive, the exponential factor in Eq. (6) will decrease with time and, as a result, oscillation will eventually die out. However, if this quantity is negative, the exponential factor increases with time and oscillation builds up in amplitude. In the critical case, $(r/L + 1/pC)$ is zero, and the exponential factor is then equal to unity. In this case, the amplitude of oscillation will remain constant. Obviously, $(r/L + 1/pC)$ cannot be negative or zero unless either r or p is negative. In the critical case when $(r/L + 1/pC)$ is equal to zero, $p = -L/rC$ and the amplitude of oscillations is constant, the expression for frequency, Eq. (7), reduces to:

$$f = \frac{1}{2\pi}\sqrt{\left(\frac{r+p}{p}\right)\left(\frac{1}{LC}\right)} \qquad (8)$$

In practice, r will be extremely small compared to p and the frequency is practically equal to $1/2\pi\sqrt{LC}$ as given by Eq. (4).

Eq. (6) shows that sustained oscillations in the circuit of Fig. 2 will occur if the resistance is negative and of such a magnitude that $p \geqq -L/rC$. Theoretically, if p is larger than L/rC, the oscillations will continuously build up in amplitude. In actual practice, two factors prevent this from happening: (1) circuit losses due to radiation will tend to increase the effective value of r until $p = -L/rC$, and the oscillation amplitude will approach a constant; (2) self-regulating factors in the device producing the negative resistance (such as those due to grid resistor bias or plate-current saturation) tend to reduce the value of p until it is equal to $-L/rC$.

In vacuum-tube oscillators, the negative resistance p is produced by the tube. The electron tube can produce an equivalent negative resistance across the tuned circuit in two ways: (1) by operation of the tube in such a way that either the screen-grid current vs. screen-grid voltage curve or the plate-current vs. plate-voltage curve has a negative slope; and (2), by feeding back a portion of the energy in the output circuit into the input circuit of the tube. Although the oscillators in both (1) and (2) actually operate as negative-resistance devices, those employing the negative-resistance characteristics of the plate or screen-grid-voltage characteristics are generally referred to as negative-resistance oscillators, while those utilizing feedback from the output circuit are generally referred to as feedback oscillators.

Eq. (6) shows that the circuit in Fig. 2 is oscillatory if p is negative and equal in magnitude to L/rC. Therefore, for the continuously oscillating case, another form of equivalent circuit for Fig. 2 may be drawn, as shown in Fig. 3.

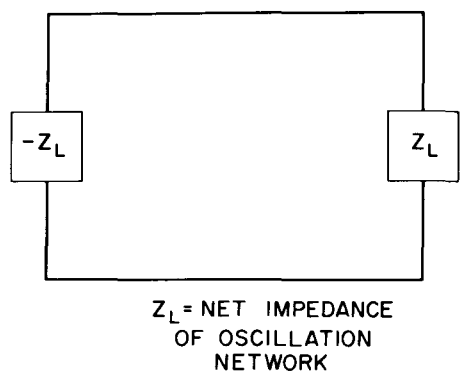


Figure 3. Another Form of Equivalent Circuit for Figure 2

Consider Z_L in Fig. 3 and p in Fig. 2 as either negative resistance or as an ac generator delivering power to replace the power losses in the oscillating circuit. Fig. 3 may be used for an analysis of practically any oscillator circuit by application of the following theorem:

> An oscillatory circuit containing a generator (or negative resistance) will have zero impedance to a test generator connected in series and zero admittance to a shunt-connected test generator.

An analysis of the circuit of Fig. 3 by means of the above theorem may be made by insertion of a test generator in shunt with the circuit as shown in Fig. 4. For oscillation to occur, the admittance Y_t must be zero.

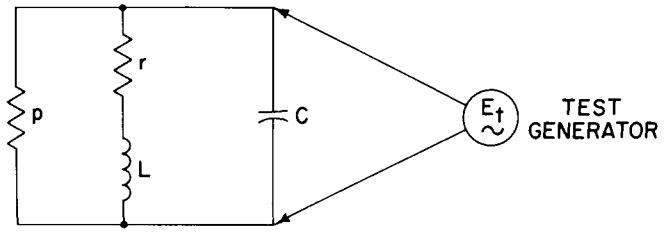


Figure 4. Test Generator Inserted in Shunt with Circuit of Figure 3

$$Y_t = \frac{1}{p} + \frac{1}{r + j\omega L} + j\omega C$$

$$= \frac{1}{p} + \frac{r}{r^2 + \omega^2 L^2} - \frac{j\omega L}{r^2 + \omega^2 L^2} + j\omega C$$

When both real and imaginary terms of the above equation are set simultaneously equal to zero, the following is obtained:

$$p = -\frac{L}{rC}$$

where the value L/rC is the resonant impedance of the circuit.

Solving for frequency gives:

$$f = \frac{1}{2\pi}\sqrt{\left(\frac{r + p}{p}\right)\left(\frac{1}{LC}\right)}$$

as in Eq. (8)

The same expression may also be obtained by substituting a series generator in the circuit and setting the series impedance equal to zero.

Other Requirements for Oscillation. A generalized equivalent circuit of an oscillator using a vacuum tube to produce the negative resistance is shown in Fig. 5.

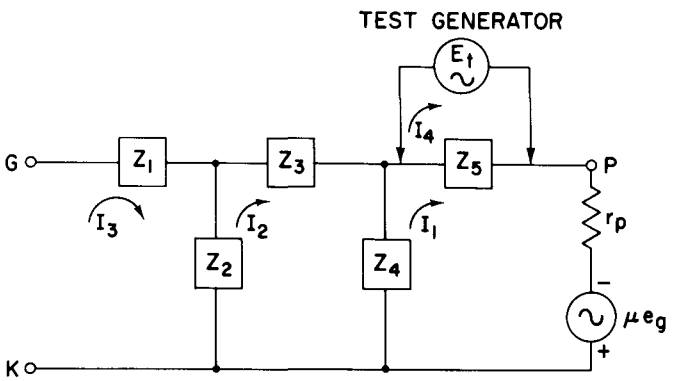


Figure 5. Generalized Equivalent Circuit of an Oscillator Using an Electron Tube to Produce Negative Resistance

For purposes of analysis, it is assumed that the grid does not go positive; therefore, $I_3 = 0$, and for oscillation to occur:

$$Y_t = \frac{I_4}{E_t} = 0$$

A solution of the above equation in terms of the impedances shown in Fig. 5 indicates that for oscillation to occur:

$$(Z_2 + Z_3 + Z_4)(r_p + Z_5) + Z_2Z_4(\mu + 1) + Z_3Z_4 = 0 \quad (9)$$

If it is assumed that the resistive components of the impedances shown in Fig. 5 are small enough to be negligible, some additional interesting facts may be obtained regarding the conditions necessary for oscillation to occur.

Substitution of $Z_1 = jX_1$, $Z_2 = jX_2$ . . ., etc., in Eq. (9) gives the following result:

$$-X_5(X_2 + X_3 + X_4) + jr_p(X_2 + X_3 + X_4) - X_2X_4(\mu + 1) - X_3X_4 = 0$$

Equating real and imaginary quantities to zero gives:

$$-X_5(X_2 + X_3 + X_4) - X_2X_4(\mu + 1) - X_3X_4 = 0$$

$$r_p(X_2 + X_3 + X_4) = 0$$

Since r_p cannot be zero, $X_2 + X_3 + X_4 = 0$

therefore, $-X_2X_4(\mu + 1) = X_3X_4$

and $-X_2(\mu + 1) = X_3$

Substituting $-X_2(\mu + 1)$ for X_3 in the expression

$$X_2 + X_3 + X_4 = 0$$

gives $\mu X_2 = X_4$

Therefore, for oscillation to occur, X_2 and X_4 must be of the same sign, i.e., Z_2 and Z_4 both must be inductances or capacitors. Similarly, since $X_2 + X_3 + X_4 = 0$, then X_2, X_3 and X_4 cannot all be of the same sign. Therefore, X_2 and X_3 must be of opposite sign and if one is an inductance, the other must be a capacitor.

Example: Consider the circuit of an oscillator using a tickler feedback coil, as shown in Fig. 6. Assume the tickler winding (L_2) is connected in the wrong polarity so that the equivalent T of the circuit in Fig. 6 is as shown in Fig. 7. In this case, the circuit cannot oscillate because Z_2 and Z_4 are opposite in sign. However, if the tickler winding is reversed, the equivalent T of the circuit becomes as shown in Fig. 8. Oscillation can now take place because Z_2 and Z_4 are of the same sign and Z_2 and Z_3 are opposite in sign. A similar analysis can be made of other oscillator circuits.

Oscillator Analysis by Feedback Amplifier Theory

It was stated earlier that all oscillators operate as negative-resistance devices. However, it often is convenient to analyze certain oscillators by feedback amplifier theory.

Let: A_f = gain of feedback amplifier
A = gain of amplifier without feedback
β = fractional portion of output voltage fed back into grid circuit
then, $A_f = A/1 - A\beta$
and if $1 - A = 0$, then $A_f = \infty$

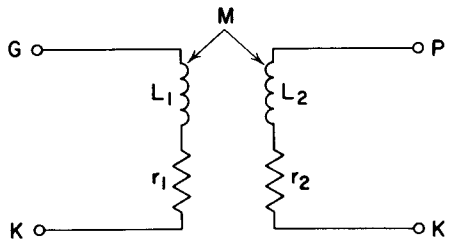


Figure 6. Circuit for Oscillator Using Tickler Feedback Coil

In other words, the circuit will have an ac output voltage with no ac input voltage, and the circuit will oscillate. A feedback amplifier and its equivalent circuit are shown in Fig. 9.

The expressions for current, voltage, amplification, and feedback factor are:

$$I_p = \frac{\mu E_g}{r_p + Z_5 + \frac{Z_4 (Z_2 + Z_3)}{Z_2 + Z_3 + Z_4}}$$

$$E_o = \frac{-\mu E_g Z_4 (Z_2 + Z_3)}{(r_p + Z_5)(Z_2 + Z_3 + Z_4) + Z_4 (Z_2 + Z_3)}$$

$$A = \frac{E_o}{E_g} = \frac{-\mu Z_4 (Z_2 + Z_3)}{(r_p + Z_5)(Z_2 + Z_3 + Z_4) + Z_4 (Z_2 + Z_3)}$$

$$A\beta = \frac{-\mu Z_2 Z_4}{(r_p + Z_5)(Z_2 + Z_3 + Z_4) + Z_4 (Z_2 + Z_3)}$$

$$1 - A\beta = 1 + \frac{\mu Z_2 Z_4}{(r_p + Z_5)(Z_2 + Z_3 + Z_4) + Z_4 (Z_2 + Z_3)}$$

If feedback is such that oscillations occur, i.e., $1 - A\beta = 0$, the above expression may then be equated to zero to give the following expression, which defines the requirements for oscillation.

$$(Z_2 + Z_3 + Z_4)(r_p + Z_5) + Z_2 Z_4 (\mu + 1) + Z_3 Z_4 = 0$$

This expression is identical with Eq. (9), obtained by using a test generator and setting either the shunt admittance or series impedance equal to zero.

Negative-Resistance Oscillators

The dynatron is the best known of the negative-resistance oscillators. This oscillator makes use of the negative slope of the plate-current characteristic produced by secondary emission (see Fig. 10 and Fig. 11).

Although the dynatron oscillator is not a particularly reliable device because it depends on a screen-grid emission factor which varies considerably during the life of the tube used, it has found some interesting uses. The frequency stability of the dynatron oscillator approaches that of a crystal oscillator without temperature control. For example, it has been used as a heterodyne wave meter.[1] Such a wave meter can be used at frequencies much higher than the fundamental frequency of the dynatron oscillator if headphones are connected in the screen-grid circuit. This makes it possible to hear beats between the harmonics of the dynatron oscillator and other oscillators. The dynatron oscillator can also be used for measuring the high-frequency resistance of a tuned circuit.[2] In this application, the negative resistance of the tube is regulated by varying the control-grid bias voltage until oscillations just begin. The negative resistance of the tube is then equal to the impedance of the parallel-resonant circuit. Thus,

$$r_p = \frac{L}{RC}$$

where L, C, and R are the inductance, capacitance, and resistance, respectively, of the tuned circuit.

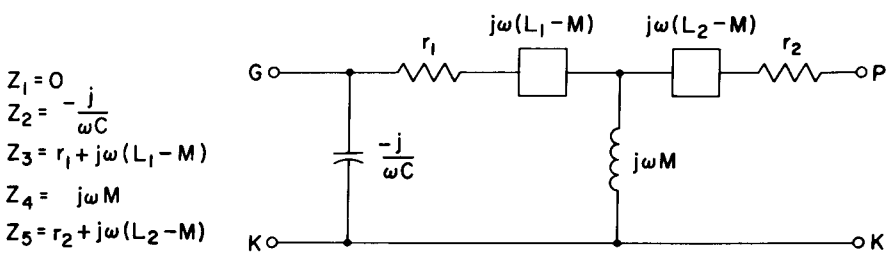


Figure 7. Equivalent Circuit for Figure 6 With Tickler Winding Connected in Wrong Polarity

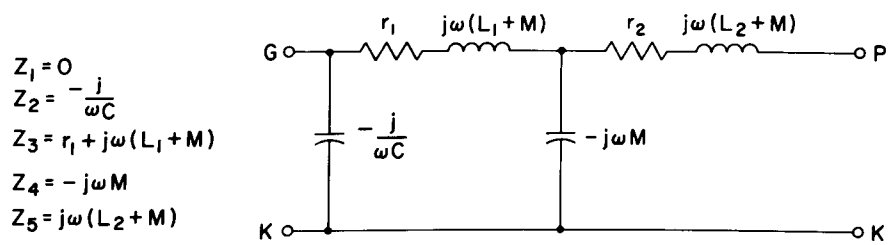
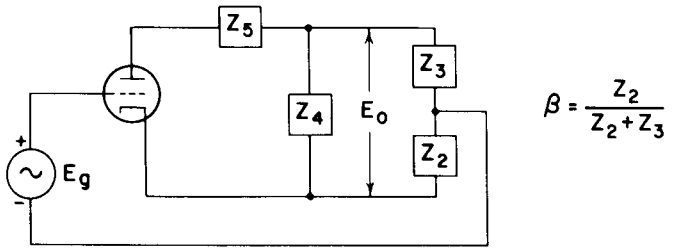


Figure 8. Circuit from Figure 7 Having Tickler Winding Connected in Proper Polarity



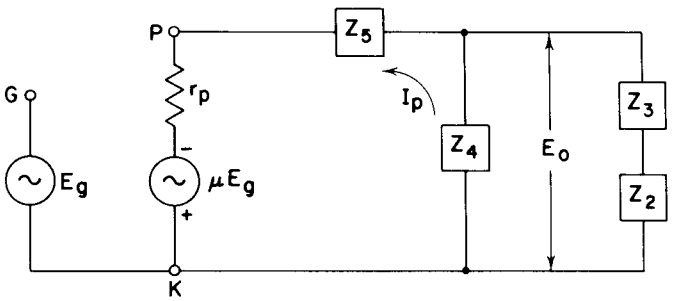


Figure 9. (A) Feedback amplifier, (B) Equivalent circuit with feedback tap omitted

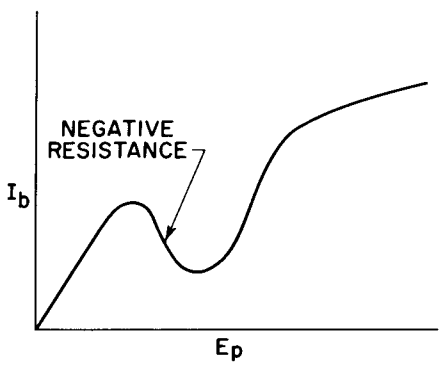


Figure 10. Curve Showing Negative Slope in Plate Current Characteristic

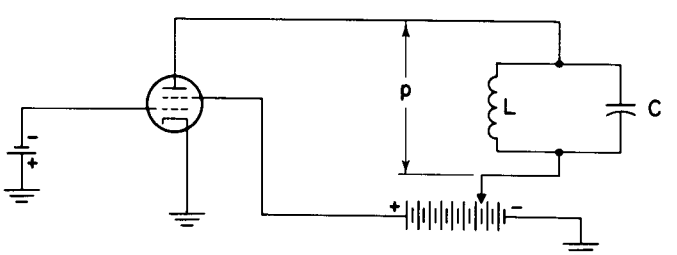


Figure 11. Circuit Diagram of Dynatron Oscillator

If the screen-grid of a pentode tube is coupled to the suppressor-grid in such a way that a change in screen-grid voltage is accompanied by a corresponding or proportioned change in suppressor-grid voltage (the suppressor-grid remaining negative and the plate positive), a portion of the screen-grid-current characteristic will have a negative slope as shown in Fig. 12.

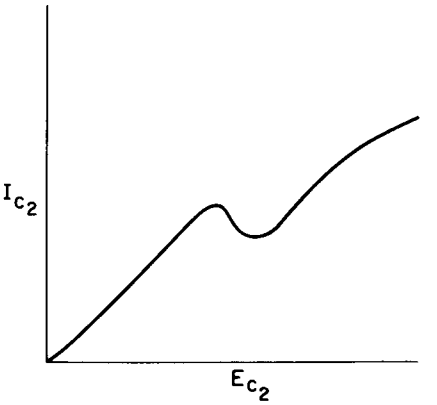


Figure 12. Negative Slope in Screen-Current Characteristic

This characteristic is utilized in the "transitron" or negative-transconductance oscillator[3] shown in Fig. 13.

A third type of negative-resistance oscillator is known as the kallirotron.[4] The basic form of this oscillator is shown in Fig. 14. The resistance between points A and B can be made negative, as shown by the variation of current flowing into B as a result of the application of a voltage between points A and B. (see Fig. 15.) A negative-resistance oscillator can be made by connecting a parallel-resonant circuit between points A and B. A practical form of such an oscillator is shown in Fig. 16.

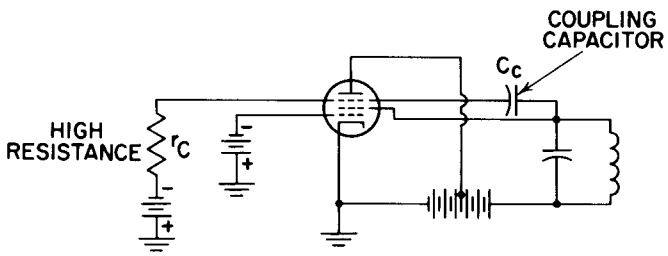


Figure 13. Circuit for Transitron, A Negative-Transconductance Oscillator

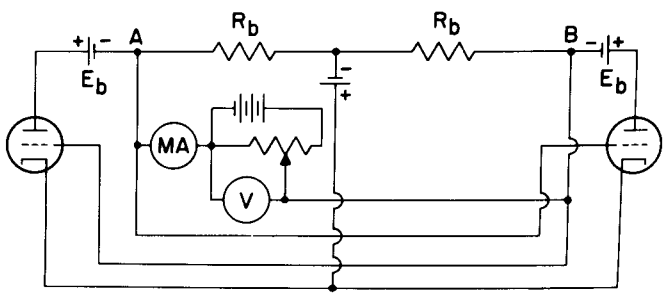


Figure 14. Basic Circuit for Kallirotron Oscillator

Feedback Oscillators

The feedback type of oscillator is almost always used as the local oscillator for superheterodyne radio or television receivers. Several typical oscillator circuits are shown in Fig. 17.

The circuits used most frequently are the Hartley oscillator, Fig. 17A and 17B; the Colpitts oscillator,

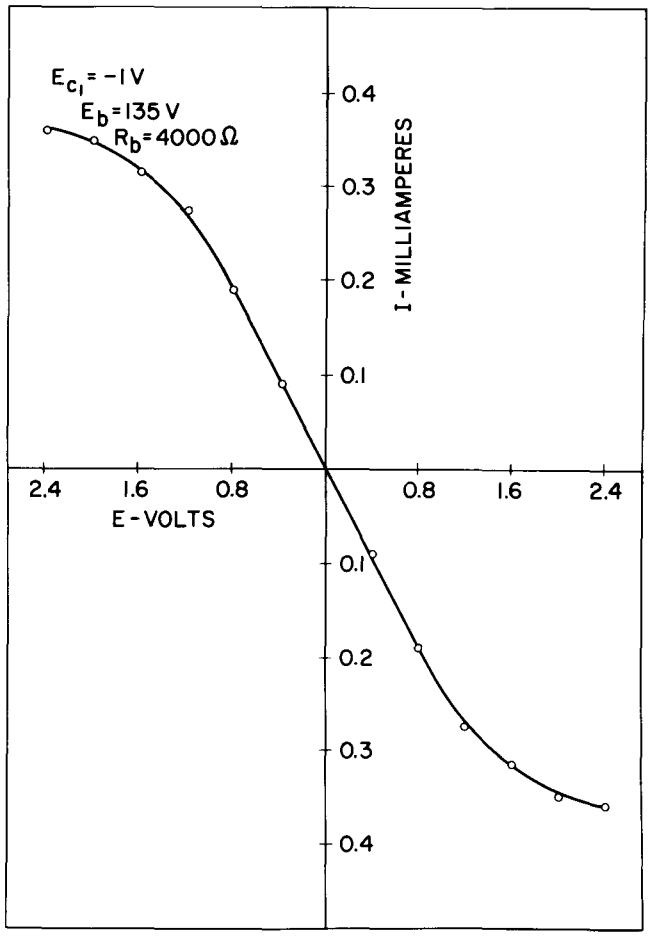


Figure 15. Negative-Resistance Characteristic of Circuit in Figure 14, Using Type 6N7 Twin Triode

Fig. 17C; and the tickler feedback oscillator, Fig. 17E. In the Hartley and Colpitts circuits, feedback voltage of the proper magnitude and phase to sustain oscillations is obtained by connecting the grid and plate of the oscillator to opposite ends of the tuned (tank) circuit. The ratio of grid-excitation-voltage to output voltage is determined by the ratio of the reactances on each side of the cathode connection (inductive reactances for the Hartley oscillator, and capacitive reactance for the Colpitts oscillator). In the tuned-grid, tuned-plate circuit, Fig. 17G, both grid and plate circuits are tuned to present an inductive reactance at the frequency of oscillation. The equivalent circuit may be redrawn, therefore, as shown in Fig. 17H, as one form of Hartley oscillator. The equivalent inductance Lp and L_G shown in Fig. 17H, are functions of the relative amounts of detuning of the grid and plate circuits. In the tickler-feedback circuits (Fig. 17E and 17F), the grid-excitation voltage is obtained by inductive coupling of the grid and plate coils. In the Meissner circuit, Fig. 17D, the tank circuit is inductively coupled to the tube, rather than directly coupled, as in the other circuits shown.

Electron-Coupled Oscillator

The electron-coupled oscillator is of considerable interest, although it finds its principal use as a generator of radio-frequency power for transmitters, rather than as a local oscillator for radio or television receivers. This circuit (Fig. 18) combines the functions of an oscillator and buffer amplifier and uses a tetrode or pentode, with the screen grid serving as the oscillator plate. Enough electrons are intercepted by the screen grid so that oscillations are maintained, but the greater portion passes between the screen-grid wires to the plate. The plate current, therefore, is modulated at the oscillator frequency, and useful power output can be obtained from the plate circuit by a tuned circuit connected to the plate which is resonant at the oscillator frequency or at some appropriate harmonic.

Grid-Circuit Considerations

Fixed-bias voltage is rarely used in oscillator circuits, except in those having very low output. Generally speaking, oscillators operate the tube under class C conditions. If a high bias voltage were supplied from a fixed supply, no plate current could flow when plate voltage was first applied and, therefore, oscillations would not start. The bias voltage must be low enough initially to allow fairly high plate current to flow. After oscillations have started, however, the bias voltage must be high enough to prevent grid or plate currents from becoming excessive and to limit the plate-current conduction angle to that required for high efficiency. The requirements of low initial bias voltage and relatively high operating bias voltage are usually met by the use of a grid resistor (commonly called grid-leak bias) to supply the bias voltage.

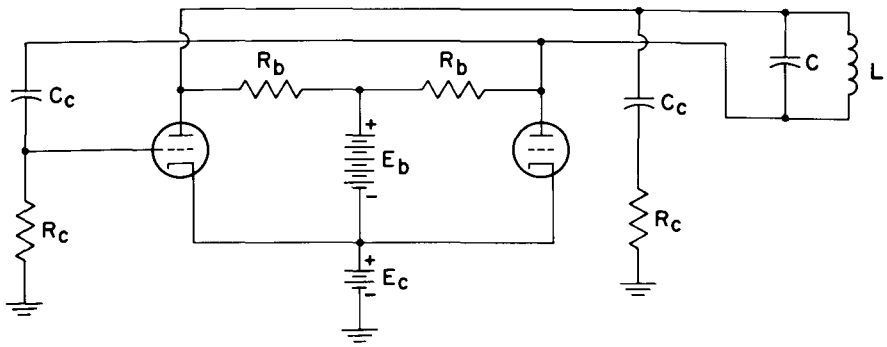


Figure 16. Practical Form of the Kallirotron Oscillator

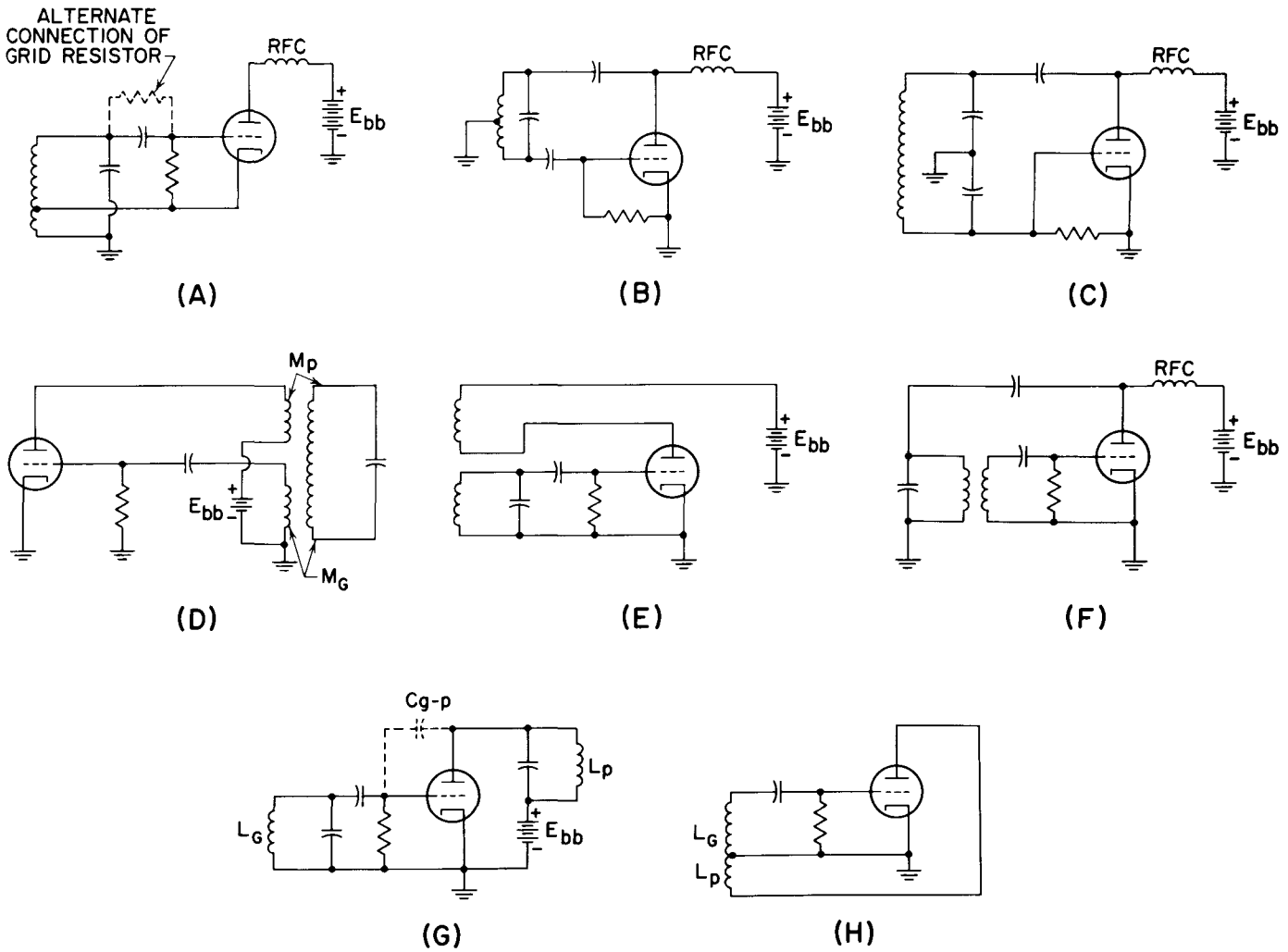


Figure 17. Typical Feedback Oscillators: (A) Hartley, series feed; (B) Hartley, shunt feed; (C) Colpitts circuit; (D) Meissner circuit; (E) Tickler feedback, tuned-plate circuit; (F) Tickler feedback, tuned-plate circuit; (G) Tuned-grid, tuned plate circuit; (H) Equivalent circuit of tuned-grid, tuned-plate oscillator

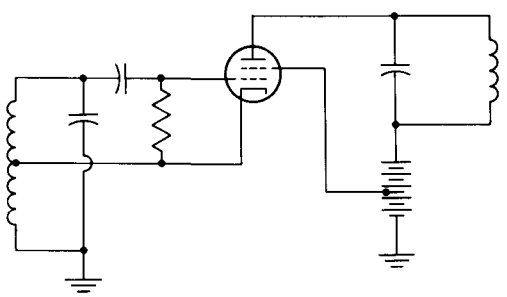


Figure 18. Electron-Coupled Oscillator Circuit Using Tetrode

In the circuits shown in Fig. 17, a resistor is connected between grid and cathode. In some cases, the grid resistor may be connected directly in shunt with the grid-blocking capacitor as shown by the alternate connection in Fig. 17A. The dc grid current flowing through this resistor produces a voltage drop which makes the grid negative with respect to the cathode. At the instant when plate voltage is first applied, there is no ac voltage applied to the grid and, consequently, no grid current. As a result, the grid bias voltage is zero, and (until oscillations begin) plate current is limited only by the current capabilities of the tube and power supply. Because the tube is operating with zero bias voltage at this instant, transconductance is very high and oscillations start. Once oscillations commence, an ac voltage is applied to the grid, grid current flows, and a negative voltage is developed across the grid resistor.

The use of the grid-resistor method not only makes the oscillator self-starting, but also contributes to the stability of operation under the desired voltage and current conditions. If the amplitude of oscillations tends to increase, the ac voltage applied to the grid will also increase, resulting in an increased grid current and higher developed grid bias voltage. The higher developed grid bias voltage tends to reduce the amplitude of oscillations. If the amplitude of oscillations decreases, the ac voltage applied to the grid is reduced, grid current and bias voltage are reduced, more plate current is allowed to flow, and the amplitude of oscillations increases.

Some form of grid-blocking capacitor is always associated with the grid resistor. When connected in the alternate connection shown in Fig. 17A, this capacitor bypasses rf currents around the grid resistor. In all cases, the capacitor forces the dc grid current to flow through the grid resistor and acts as a filter to maintain some average value of grid bias voltage.

Intermittent Operation. If the time constant of the grid-resistor/grid-capacitor combination is too large, the bias voltage developed across the grid resistor cannot adjust itself rapidly to sudden changes in the amplitude of oscillation. This condition results in an effect known as intermittent operation, "squegging", or super-regeneration. When the rate of change of bias voltage is so slow that oscillation dies out before the bias voltage can change appreciably, then the oscillator becomes very unstable. An explanation of this effect follows.

When plate voltage is first applied, oscillations start and build up rapidly to a high amplitude. A high ac voltage is then applied to the grid, causing a large grid current to flow and a large value of grid bias voltage to be developed. This bias voltage decreases the amplitude of oscillations. If the time constant of the grid-resistor/grid-capacitor combination is so high that the capacitor cannot discharge rapidly, the grid bias voltage remains at a high value when the amplitude of oscillations decreases. Oscillation then dies out completely, and does not start again until the grid capacitor discharges sufficiently to reduce the grid bias voltage and produce transconductance high enough to start oscillations again. This process then repeats itself at a rate of about (1/RgCg) times per second, determined by the RC time constant in the grid circuit. Intermittent operation may also result when the feedback voltage is barely sufficient to produce oscillation, when tube emission or transconductance is low, when feedback is excessively high, or when the oscillator is loaded too heavily.

To minimize trouble due to intermittent operation, it is important to keep the value of grid capacitor as small as practical. The grid resistor value is usually fixed by grid-bias voltage requirements. The value of grid capacitor is especially critical when a high-mu tube is used as an oscillator, because a small change in grid bias voltage produces a large change in the amplitude of oscillation. Feedback should also be adjusted carefully to minimize trouble due to excessive or insufficient feedback.

Blocking. If the grid current should be reversed in some way, the developed bias voltage would then make the grid positive with respect to the cathode. A reversal of grid current may result from secondary or thermal emission from the grid. Secondary emission may be caused by a transient of high voltage applied to the grid, or by a high ac voltage developed because of excessive feedback. Thermal emission may be caused by excessive grid-driving power (excessive feedback). If the grid resistor value is quite high, the grid may become so positive that the reverse grid current is maintained and the tube has a constant positive bias voltage. This condition usually results in such high plate dissipation that the tube is destroyed unless plate voltage is removed quickly.

To reduce the tendency toward blocking, it is desirable that the grid resistor be no larger than required for proper operation of the oscillator, and that grid drive be such that grid temperature does not become excessive.

OSCILLATOR CIRCUITS FOR SPECIAL APPLICATIONS

Crystal Oscillators

The frequency stability of an oscillator can be made exceptionally high if the usual tank circuit is replaced by a quartz crystal having a mechanical resonant frequency equal to the desired frequency of oscillation. Quartz, like a number of other crystalline substances, has the ability to transform mechanical energy into electrical energy, and vice versa. This property is known as the piezo-electric effect. Piezoelectric crystals have natural frequencies of vibration ranging from a few thousand cycles per second to several megacycles per second. The natural frequency of vibration is a function of the kind of material, the way in which the plate is cut from the natural crystal, and the dimensions of the plate. Because of the piezoelectric effect, the crystal can be substituted for the resonant circuit. A crystal and holder and its equivalent circuit are shown in Fig. 19.

Many oscillator circuits use quartz crystals as the frequency-controlling element; the circuit most commonly used is shown in Fig. 20.

If the crystal in Fig. 20 is replaced by the equivalent circuit of Fig. 19, the circuit in Fig. 20 becomes a tuned-grid, tuned-plate oscillator in which the crystal serves as the tuned-grid circuit.

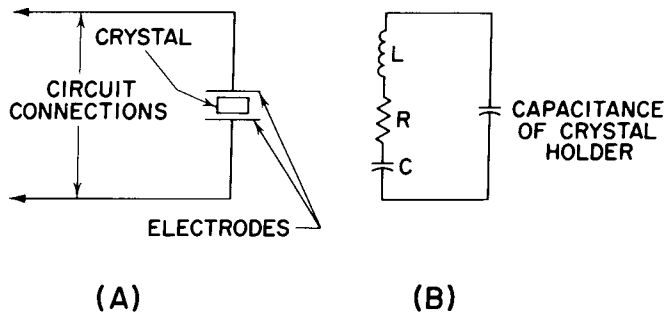


Figure 19. (A) Crystal and holder, (B) Equivalent circuit

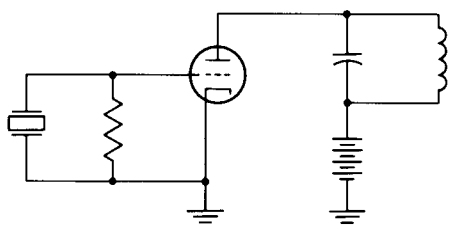


Figure 20. Commonly Used Oscillator Circuit Using Crystal as Frequency-Control Element

Magnetostriction Oscillator

Nearly all mechanical vibrators exhibit a sharpness of resonance and frequency stability approaching that of the quartz crystal. These features are utilized in another type of oscillator known as the magnetostriction oscillator.[5] The magnetostriction oscillator makes use of an unusual characteristic of several metals, such as nickel, iron-nickel alloys, and Nichrome. These metals exhibit a slight change in physical dimensions when their degree of magnetization is changed. A typical circuit of a magnetostriction oscillator is shown in Fig. 21.

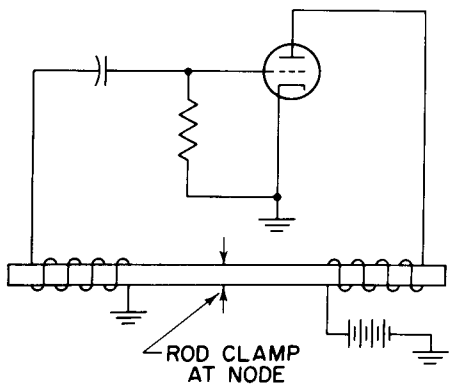


Figure 21. Typical Circuit of Magnetostriction Oscillator

The nickel rod, which passes through the grid and plate coils, is clamped at the mechanical node and is magnetized by the flux produced by plate current through the plate coil. Mechanical deformation of the rod due to the magnetic flux produced by the plate coil produces a change in the flux linking the grid coil. This change induces a grid voltage which, in turn, varies the plate current and produces further deformation of the rod. The amplitude of the mechanical motion of the rod builds up at a frequency dependent upon the physical dimensions of the rod. Fixed grid and plate coils can be used over a considerable range of frequencies. Different frequencies are obtained by the substitution of rods of different alloys or dimensions.

The magnetostriction oscillator does not compete with the crystal oscillator because its use is confined to sonic and supersonic applications, from 1 to 100 Kc. It finds application in such fields as biological work, supersonic cutting and drilling tools, and the like.

Low-Frequency Oscillators

Theoretically, any of the previously described oscillators may be used at audio frequencies, as well as at radio and supersonic frequencies. However, such limitations as wide frequency-range requirements and the weight, size, and cost of audio-frequency inductors have led to the development of special oscillator circuits for the generation of frequencies of the order of 5 to 500,000 cycles per second.

Beat-Frequency Oscillator. A block diagram of a beat-frequency oscillator for producing audio frequencies is shown in Fig. 22.

The beat-frequency oscillator is actually made up of two rf oscillators. One oscillator operates at a fixed frequency, such as 1,000,000 cycles; the other oscillator is variable from 1,000,000 to 1,020,000 cycles, for example. The two rf signals are mixed in the mixer circuit and the difference frequency appears in its output. Since, in addition to the desired frequency, the mixer circuit produces the sum frequency, plus all the harmonics and combinations of the radio frequencies, an rf filter must be added to remove the undesired signals.

The two rf oscillators must be carefully shielded and decoupled from each other to minimize the tendency to synchronize or pull in to the same frequency when the difference frequency becomes small. The principal advantage of the beat-frequency oscillator arises from the fact that a continuous output from a few cycles per second throughout the entire audio range may be obtained by a single turn of a dial controlling a variable capacitor.

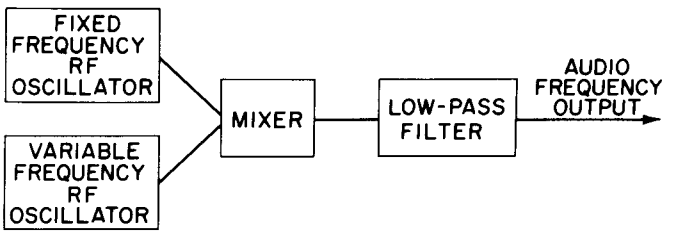


Figure 22. Block Diagram of Beat-Frequency Oscillator in Audio Frequency Range

RC Oscillators. An RC oscillator is one which uses only resistance and capacitance in the frequency-determining network. A practical form of one type of RC oscillator which has been successfully used in a commercial audio-signal generator is shown in Fig. 23.

The above circuit is that of a feedback amplifier with $1-A\beta = 0$. An analysis of the circuit by feedback amplifier theory, as discussed earlier, yields the following expression for the frequency of oscillation.

$$f = \frac{1}{2\pi\sqrt{R_1R_2C_1C_2}} \qquad (10)$$

If the circuit constants are chosen such that $R_1 = R_2$ and $C_1 = C_2$, then Eq. (10) becomes

$$f = \frac{1}{2\pi R_1C_1} \qquad (11)$$

Usually this type of oscillator is constructed with C_1 and C_2 on a common tuning shaft. As the tuning shaft is rotated, C_1 and C_2 change at the same rate and, therefore, the frequency of oscillation will be inversely proportional to the capacitance of C_1 or C_2. As a result such an oscillator has a much wider tuning range than any of the oscillators previously discussed (excepting the beat-frequency oscillator) which has its frequencies of oscillation inversely proportional to the square root of the tuning capacitance.

With the RC oscillator, conventional variable capacitors can be used to obtain a tuning range of 10:1 and

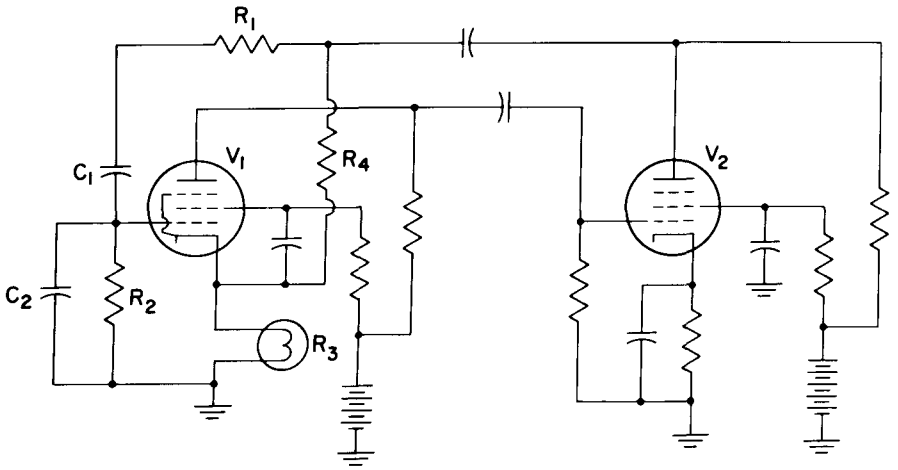


Figure 23. Practical Form of RC-Oscillator Circuit

variation of resistors R_1 and R_2 in steps of decimal values can be used to give decimal multiplying factors.

The incandescent lamp in the cathode of V_1, together with resistor R_4, provides inverse feedback which stabilizes the output voltage of the oscillator and reduces distortion under wide variations of supply voltage, tube characteristics, and frequency. An increase in the amplitude of oscillation increases the current through the lamp, increasing its resistance and thus increasing the feedback. A decrease in amplitude tends to decrease the lamp resistance and thus decrease the feedback.

One disadvantage of the circuit shown in Fig. 23 is that C_1 must be isolated from ground. This requirement dictates the use of a ganged capacitor with the two sections insulated from each other and at very low frequencies, this insulation must have extremely high leakage resistance.

A circuit which places both rotor sections of the tuning capacitor at ground potential is shown in Fig. 24.[6] In this circuit, the frequency of oscillation is given by:

$$f = \frac{1}{2\pi\sqrt{R_1R_2C_1C_2}}$$

or, if $R_1 = R_2$ and $C_1 = C_2$,

$$f = \frac{1}{2\pi R_1C_1}$$

as in the circuit shown in Fig. 23.

Stabilization of output and reduction of distortion can be obtained, as in Fig. 23, by connecting an incandescent lamp in series with, or in place of R_3. Stabilization may also be provided in both circuits by the use of amplified and delayed automatic gain control (agc). The output of the oscillator is sampled, amplified, rectified, and filtered. It is then fed to the grid of the first tube as a negative bias voltage which tends to vary in a direction to compensate for any variation in output of the oscillator.

Oscillators Using Resonant Lines. The circuits previously discussed have used "lumped constant" elements (inductors and capacitors) for the resonant circuits. Such circuits can be used with conventional miniature tubes through the vhf television band (250 Mc). However, for the uhf television band (470 to 890 Mc), and higher, it is virtually impossible to construct an oscillator with lumped constants. The effective tube capacitances and lead inductances become such a large proportion of the total circuit inductance and capacitance that conventional coils and capacitors cannot be constructed small enough to complete the circuit.* Tank circuits for ultra-high frequency oscillators, therefore, are always either resonant transmission lines or cavity resonators. Such circuits have very high Q** and high impedance and have relatively large dimensions in proportion to the resonant wavelength.

Although it should theoretically be possible to construct an oscillator of any of the types shown in Fig. 17 using resonant lines, the resonant-line oscillator is usually of the Colpitts type with the lines connected to the grid and plate. Normally, the only tuning capacitances are the tube capacitances. Calculations for the dimensions of the resonant line are made with the assumption that the tube represents a capacitance across one end of the line. Fig. 25 shows a typical circuit of a resonant-line oscillator, such as might be used with tube type 6AF4-A.

While it is possible to construct oscillators with open-ended transmission lines, it is the usual practice to have an rf short (in the form of a capacitor) at the end farthest from the tube. For such a configuration, the equivalent line length must be one-quarter wavelength or an odd multiple thereof. The approximate length of the transmission line may be calculated by transmission-line theory, as follows.

---

* For example, the 6AF4-A will oscillate at approximately 1200 Mc with its grid and plate terminals shorted.

** Q is a term used to indicate the relative losses in a tuned circuit. Mathematically, $Q = \omega L/r$ where ωL is the inductive reactance and r is the series resistance in the circuit.

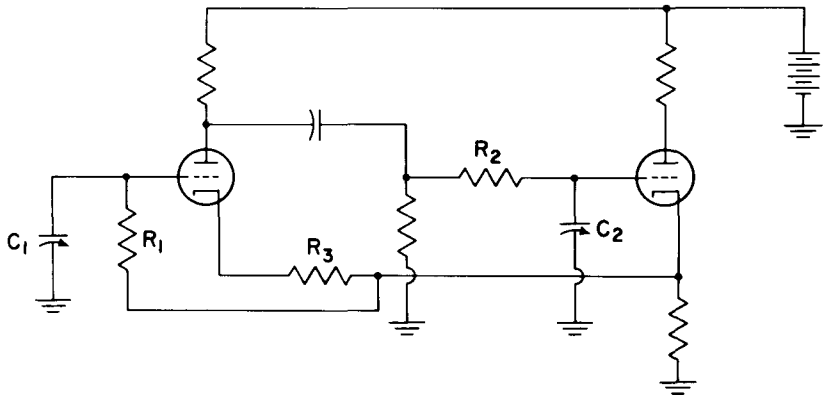


Figure 24. RC-Oscillator Circuit Having Both Rotors of Tuning Capacitor at Ground Potential

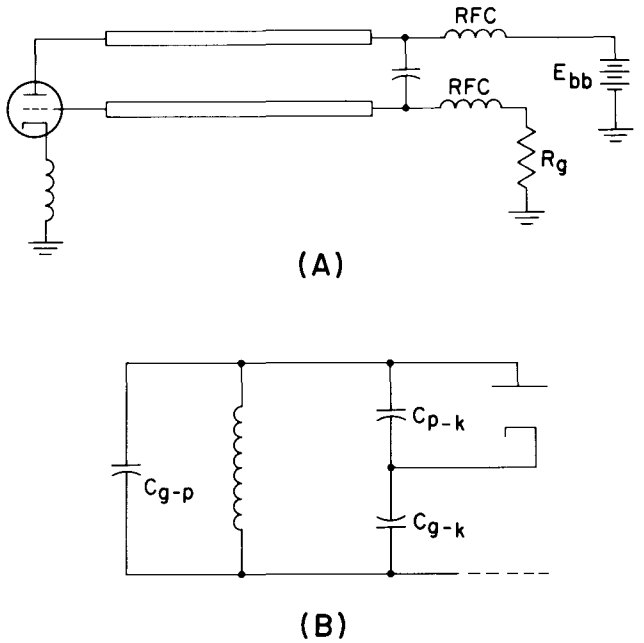


Figure 25. (A) Resonant-line oscillator; (B) Equivalent circuit of a resonant-line oscillator

The input impedance of a lossless transmission line shorted at the far end is:

$$Z_i = jZ_o \tan \frac{2\pi l}{\lambda} \qquad (12)$$

$$Z_o = \sqrt{\frac{L}{C}} \qquad (13)$$

where Z_i = input impedance of line in ohms
Z_o = characteristic impedance of line in ohms
L = inductance per unit length of line in henries per meter
C = capacitance per unit length of line in farads per meter
l = length of line in meters
λ = wavelength in meters

It is evident from Eq. (12) that the input impedance of the line is inductive if the line is less than one-quarter wave length long, capacitive from one-quarter wave length to one-half wave length, inductive from one-half wave length to three-quarter wave length, etc.

Eqs. (12) and (13) can be used to compute the proper length of transmission line required to resonate at any given frequency. In most oscillator circuits, the transmission line will have the grid-plate capacitance of the oscillator tube connected across its open end. The "line-shortening" effect of this capacitance must be considered in computing the physical length of the transmission line. For resonance, the inductive input impedance of the transmission line must be equal to the capacitive reactance of the tube grid-plate capacitance.

Computation of the transmission line length can be considerably shortened by use of the curves shown in Fig. 26. These curves are derived from Eq. (12).

It must be remembered that at ultra-high frequencies, the inductive reactance of the tube leads and the capacitive reactance of tube sockets cannot be neglected. Usually, an estimate of the equivalent length of transmission line represented by the tube leads and socket can be made. If an intelligent estimate is made of the effect of the tube and socket, the computed line length should closely approach that required for the desired frequency. The exact length will, of course, require adjustment when the oscillator has been constructed and the frequency is actually measured.

Example: Assume that tube type 6AF4-A is used in the circuit of Fig. 25 at a frequency of 500 Mc. Assuming the tramsmission line is made up of 1/8-inch rods spaced 3/8-inch apart, determine what length of line is required.

Z_o of the transmission line can be determined from tables or may be computed from Eq. (14).

$$Z_o = 276 \log_{10}\left(\frac{d-r}{r}\right) \qquad (14)$$

where d is the line spacing and r is the radius of the line, measured in the same units

From Eq. (14), Z_o = 193 ohms.

The total capacitance across the open end of the line is:

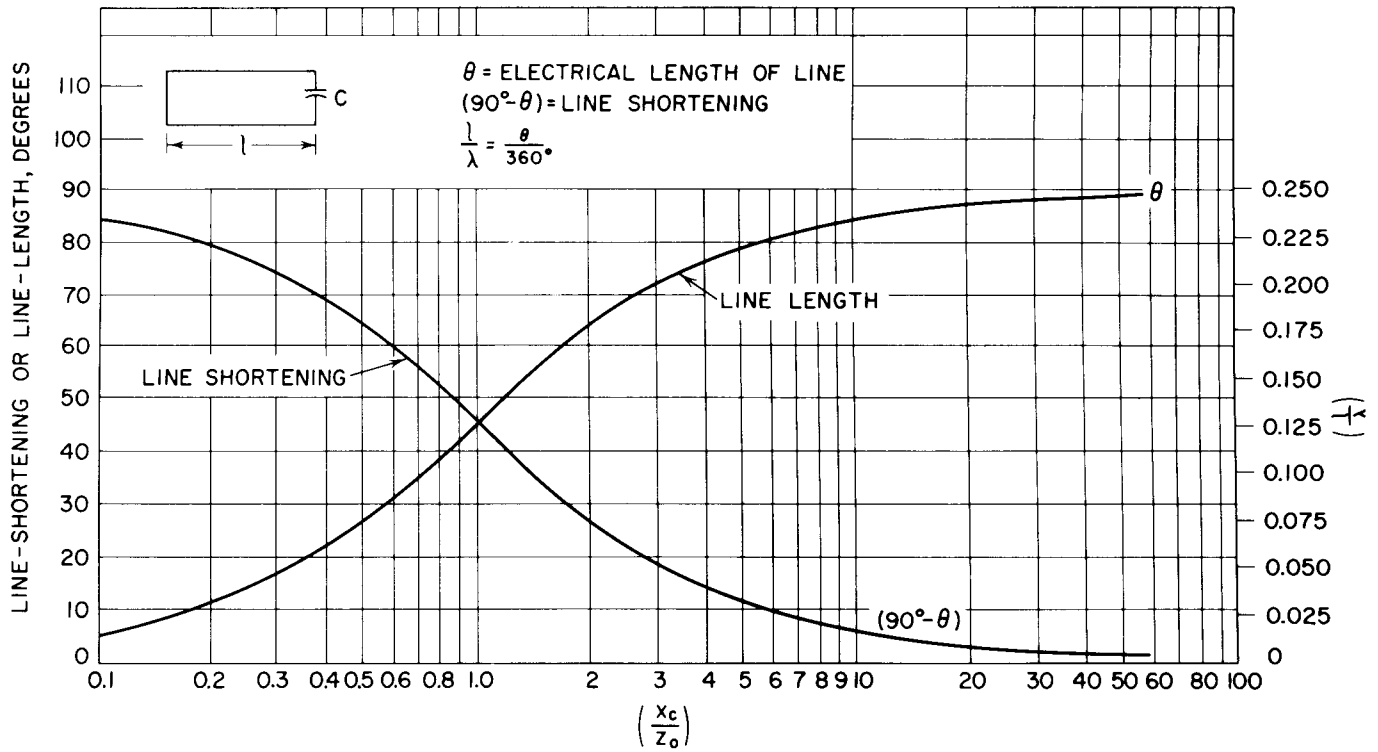


Figure 26. Curves for Computation of Transmission-Line Length

$$C = C_{g\text{-}p} + \frac{(C_{p\text{-}k})(C_{g\text{-}k})}{(C_{p\text{-}k} + C_{g\text{-}k})} = 1.9 + \frac{(0.45)(2.2)}{(0.45+2.2)} = 2.27\ \mu\mu f$$

$C_{g\text{-}p} = 1.9\ \mu\mu f$

$C_{p\text{-}k} = 0.45\ \mu\mu f$

$C_{g\text{-}k} = 2.2\ \mu\mu f$

From the curves of Fig. 26,

$$X_c = \frac{1}{2\pi(500 \times 10^6)(2.27 \times 10^{-12})} = 140 \text{ ohms}$$

$$\frac{X_c}{Z_o} = \frac{140}{193} = 0.725$$

$$\frac{l}{\lambda} = 0.1$$

$$l = (0.1)(\lambda) = (60)(0.1) = 6 \text{ cm or } 2.36 \text{ in.}$$

The estimated equivalent length of line represented by the tube and socket is approximately 0.75 inch; the actual length of external transmission line = 2.36-0.75 = 1.61 inches. In an actual circuit, it was found that the transmission-line length for 500 Mc was 1.25 inches.

Coaxial Transmission Line Oscillators. Resonant-line oscillators using miniature tubes normally make use of parallel-wire transmission lines. However, oscillators using special uhf tubes, such as the pencil and lighthouse triodes, usually are made with coaxial transmission lines. The computations for coaxial-transmission-line oscillators are similar to those discussed above, except that Eq. (15) must be used for characteristic impedance of the coaxial line.

$$Z_o = 138 \log_{10} \frac{b}{a} \qquad (15)$$

where b = inside radius of outer conductor
a = radius of center conductor (b and a must be in the same units)

Most miniature tubes will not oscillate at frequencies above 1000 Mc and special tubes must be used. However, even though the lead inductance of the lighthouse and pencil triodes is extremely small, the line-shortening effect of the tube capacitances becomes so great that the length of transmission line becomes extremely short. Accordingly, for frequencies of the order of 2000 to 3000 Mc, a special circuit, known as the reentrant oscillator, is used. A sketch of such an oscillator, using a pencil triode, is shown in Fig. 27.[7]

The electrical behavior of a reentrant oscillator is extremely complicated; a simplified explanation of its operation follows. For those who desire to analyze the

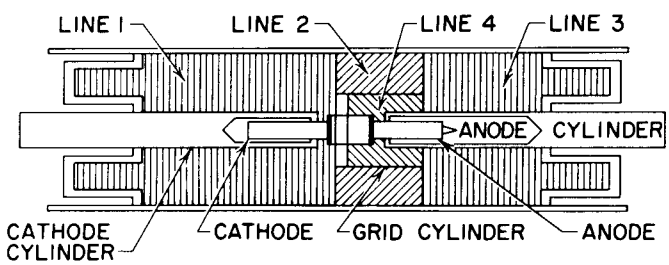


Figure 27. Sketch of Reentrant Oscillator

circuit more thoroughly, Whinnery has analyzed the re-entrant oscillator as applied to the lighthouse triode in terms of an equivalent circuit.[8]

The circuit for Fig. 27 may be considered to be made up of four transmission lines: line 1 having the cathode cylinder as its center conductor and the outer cavity wall as its outer conductor; line 2 having the grid cylinder as its inner conductor and the outer cavity wall as the outer conductor; line 3 having the anode cylinder as its inner conductor and the outer cavity wall as the outer conductor; and line 4 having the anode cylinder as its inner conductor and the grid cylinder as its outer conductor. Line 2 is the transmission line along which energy is propagated from the plate-grid region of the cavity to the cathode-grid region. The length of line 2, together with the lengths of lines 1 and 3, determines both the magnitude and phase of the feedback voltage between cathode and grid. The length of line 4 usually is the principal element determining the resonant frequency of the cavity.

If the diameter of the outer cavity cylinder is large compared to the diameter of the grid cylinder (Z_0 of line 2 high), there will be a large discontinuity at the open ends of lines 2 and 4. Then line 4 will act essentially as a half-wave line (electrical - not physical length) and the resonant frequency of the oscillator will be almost entirely dependent on its physical length. The lengths of lines 1 and 3 will affect the phase and magnitude of the feedback, but will have only a minor effect on the frequency.

If the diameter of the outer cavity cylinder is only slightly greater than the diameter of the grid cylinder (Z_0 of line 2 low), the discontinuity at the open end of lines 2 and 4 will be small. In such a case, lines 3 and 4 will have nearly the same characteristic impedance and, in fact, they may almost be considered as a single transmission line. Then the frequency, as well as the feedback, is largely determined by the length of line 3. However, since there will always be some discontinuity at the open ends of lines 3 and 4, there will always be a large portion of the stored energy in the circuit contained in line 4. Therefore, the length of line 4 will always be a large factor in determining frequency and there will be a limit to the frequency variation obtainable by adjustment of line 3 alone.

NONSINUSOIDAL OSCILLATOR CIRCUITS

A great many circuits have been developed for generating waveforms which are not sinusoidal. Only a few of the more common circuits are considered here. Such oscillators are used as sweep generators for cathode-ray oscilloscopes and television receivers, timing circuits for radar, and many other applications.

Multivibrator Circuits

In the multivibrator, the feedback voltage is so large that the tube is driven beyond cutoff. It remains cut off for a period of time determined by the RC constants in its grid circuit. Some of the more common multivibrator circuits are shown in Fig. 28.

From an examination of the circuits in Fig. 28, the multivibrator obviously is an RC-coupled amplifier with output of the second stage coupled to the grid of the first stage. In Fig. 28A and B, the output of the second stage is coupled to the grid of the first stage by means of a capacitor. In Fig. 28C, the coupling is accomplished by means of a cathode resistor common to both stages.

In the multivibrator, as well as in many other nonsinusoidal generators, the mathematical analysis used for sine-wave generators is no longer applicable. The waveforms must be calculated by a method of transient analysis, with the vacuum tube represented as a resistor in series with a switch.

The circuit in Fig. 28A operates in the following manner. When voltages are applied to the circuit, currents

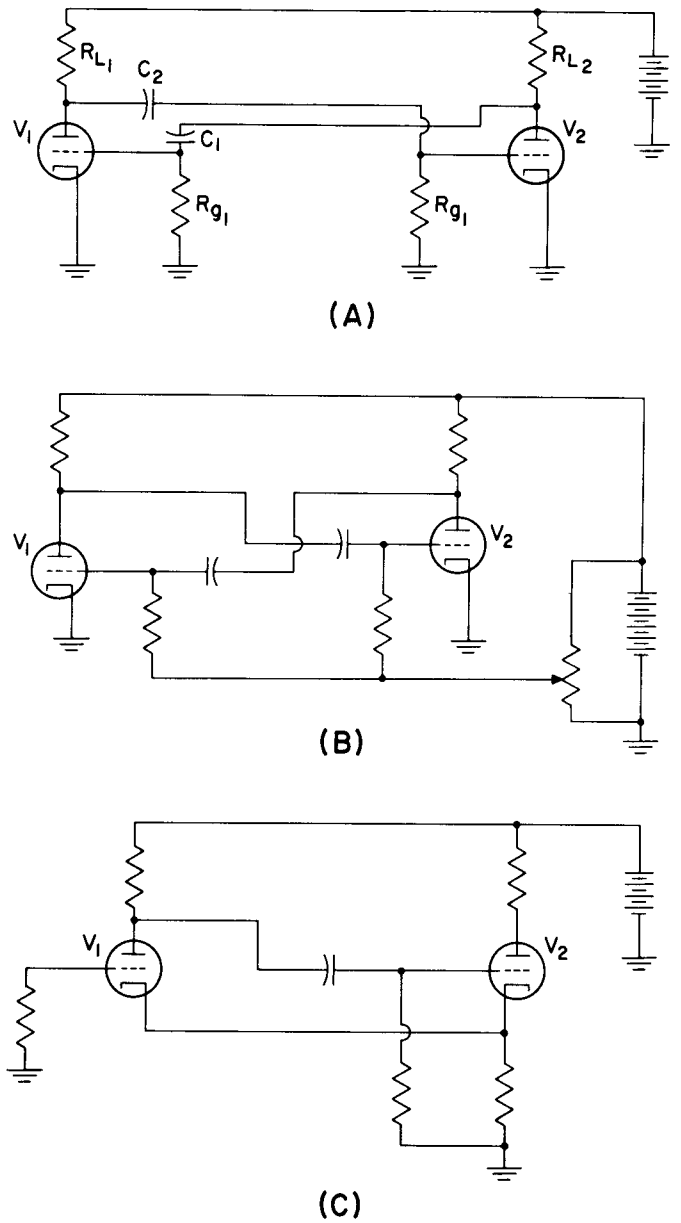


Figure 28. Common Multivibrator Circuits: (A) Plate-coupled multivibrator, (B) Biased multivibrator, (C) Cathode-coupled multivibrator

begin to flow in the plate circuits of the tubes. If the two tubes have identical characteristics and the resistors in each plate circuit are the same, the currents through the two tubes may initially be very nearly the same. However, perfect balance between the two tubes is impossible and there will always be some slight difference in the two currents. Assume that initially a slight increase in the current drawn by V_2 occurs. This slight increase causes an increase in the voltage drop across resistor R_{L2}, and a decrease in the plate voltage of V_2. Since the voltage across C_1 tends to remain constant, the decrease in plate voltage of V_2 is accompanied by a decrease in the grid voltage of V_1. This decrease in grid voltage, in turn, decreases the plate current of V_1. The decrease in plate current of V_1 results in higher voltage at the plate of V_1 due to reduced drop across R_{L1}. This increase in plate voltage causes an increase in the plate current of V_2. This action is cumulative, or regenerative, and ends with the plate current of V_1 reduced to zero and the plate current of V_2 at a maximum value. With V_1 still cut off, the charge on capacitor C_1 will leak off through resistor R_{g1} and, at some point, the grid voltage of V_1 will become such that the tube will again begin to conduct. This condition results in a negative voltage being applied to the grid of V_2 and the start of a cumulative action, as described above, which ends with V_2 cut off and V_1 conducting. The cycle then repeats. This switching action actually occurs in a very short time, in less than a microsecond in a properly designed multivibrator.

The method of transient analysis used with multivibrators will be illustrated by an example applying to Fig. 28A. Assume V_1 and V_2 are 6J5's; E_{bb} = 250, $R_{L1} = R_{L2} = 100,000$ ohms, $R_{g1} = R_{g2} = 500,000$ ohms, and $C1 = C_2 = C$. For purposes of analysis, assume that V_1 is cutoff and V_2 has been conducting a long time. From a 100,000-ohm load line drawn on the characteristic curves of the 6J5, it is evident that $e_{b2} = 30$ volts. At this time, the capacitor between the plate of V_1 and the grid of V_2 will be charged to 250 volts. At the instant conduction is transferred to V_1, the static r_p of V_1 obtained from the $e_c = 0$ characteristic and 100,000 ohm load line is approximately 13,300 ohms. The equivalent circuit of the multivibrator may then be drawn as shown in Fig. 29. The plate voltage may be expressed by the following equation:

$$\frac{eb_1}{13,300} + \frac{eb_1 - 250}{100,000} + \frac{eb_1 - 250}{500,000} = 0$$

Solving the above equation gives: $e_{b1} = 39.8$ volts

The grid voltage e_{g2} of V_2 at the instant conduction is transferred to V_1 will be: $e_{g2} = -250 + 39.8 = 210.2$ volts.

From the curves for the 6J5, it is seen that the cutoff voltage at a plate voltage of 250 volts is -15 volts. Therefore, V_2 will remain cut off until e_{g2} reaches -15 volts. The rate at which e_{g2} approaches cutoff is determined by the coupling capacitor C, the load resistance R_{L2}, plate resistance r_p, and grid resistance R_{g2}. The effective discharge resistance for C is:

$$R = 500,000 + \frac{(100,000)(13,300)}{113,300} = 511,700$$

$$t = RC \log_e \frac{210.2}{15} = (1.35 \times 10^6)C$$

The time t is the period for a half cycle; therefore, the frequency of oscillation is:

$$f = 1/2t = 1/2\,(1.35 \times 10^6)C = 3.73 \times 10^{-7}/C$$

Fig. 30 shows the relationship between frequency and coupling capacitance when the coupling capacitors are equal. When the multivibrator circuit is not symmetrical (i.e., the resistors or capacitors in the grid or plate of V_1 are different in value from those on the grid or plate of V_2) the output waveforms will not have the same period for each half cycle. It then becomes necessary to compute the waveforms for both half cycles in the same manner (as just described) and to determine the two half periods, t_1 and t_2; $t_1 + t_2$ is then the period for a full cycle. A similar method of analysis may be applied to the circuits in Fig. 28B and 28C.

The method of analysis just described applies only for relatively low frequencies where the coupling capacitors are large compared to tube interelectrode capacitances and to stray wiring capacitances. For high frequencies, these capacitances must be included in the equivalent circuit.

The frequency stability of a multivibrator such as that shown in Fig. 28A depends upon the intersection between the exponential grid voltage discharge curve, the sharpness of the cutoff voltage, and the cutoff voltage at which conduction starts. If the discharge curve is nearly horizontal as it approaches cutoff, the intersection will be poorly defined, the exact time of conduction will also be poorly defined, and the frequency stability will be poor. If the intersection is sharp, the point of conduction will be sharply defined, and the frequency stability will be improved. A sharp intersection may be obtained by returning the grids of V_1 and V_2 to a positive bias voltage, usually to the plate supply voltage, as shown in Fig. 28B; Fig. 31 illustrates how this improvement is accomplished.

Because the effective grid-cathode resistance during grid conduction is small compared with that of the grid resistor, the grid voltage is affected very little by the positive bias voltage during the part of the cycle when the grid is conducting; the plate waveforms are likewise affected very little. However, during the part of the cycle when the grid is negative, the grid voltage relaxes toward the positive bias-voltage value rather than towards zero, as shown in Fig. 31. This behavior has two important effects: by shortening the relaxation time, it increases the frequency; and by increasing the rate of change of grid voltage at the time cutoff is reached, it improves the frequency stability. A simple calculation shows that the frequency stability is improved by this behavior. Suppose that the grid resistors, in the example just discussed, had been returned to +250 volts. The half period would then be:

$$RC \log_e \frac{460.2}{265.2} = RC \log_e 1.7 = 0.53\,RC$$

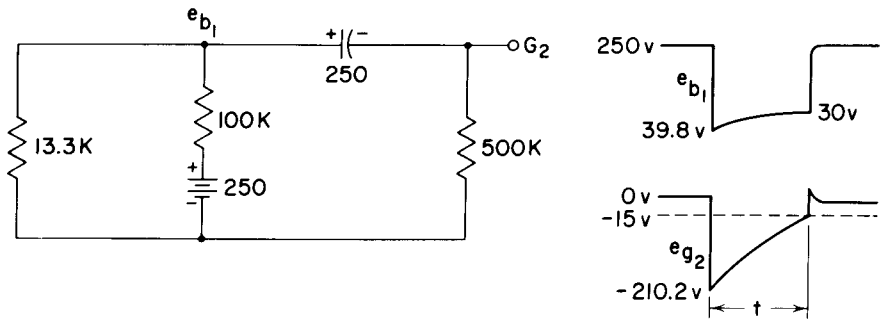


Figure 29. Equivalent Circuit and Waveforms of Multivibrator Shown in Figure 28A

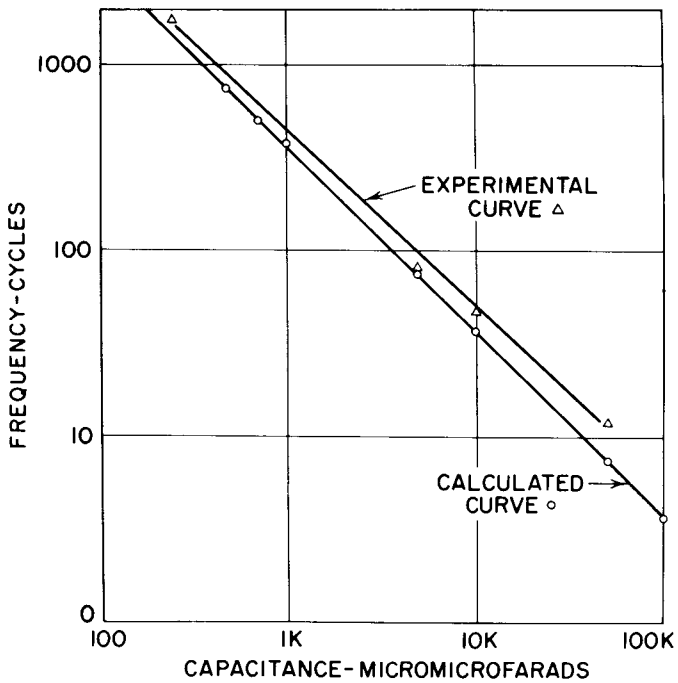


Figure 30. Symmetrical Multivibrator: Calculated and Measured Variation of Frequency with Coupling Capacitance

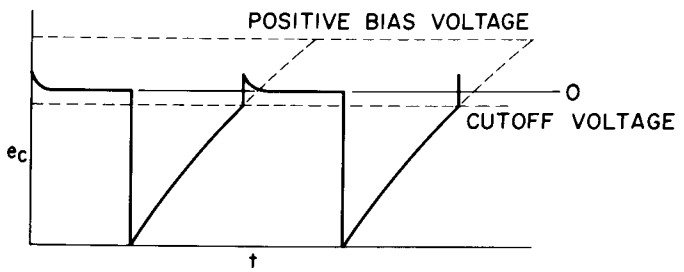


Figure 31. Grid-Voltage Waveform for Circuit in Figure 28B

If RC were 51,170 microseconds, the half period would be 27,100 microseconds. If the grids are returned to the cathode, the half period would be 135,000 microseconds, indicating that the frequency may be increased almost fivefold by returning the grid resistors to a positive voltage. Now, if one of the tubes is changed, resulting in a change of cutoff voltage from -15 to -16 volts (in the case of no bias voltage) the new half period would be about 131,000 microseconds, or a change of about three per cent. In the case of applied bias voltage, the new half period would be about 26,800 microseconds, or a change of about 1.2 per cent.

The dependency of frequency upon the bias voltage value provides a very convenient method of frequency control. By suitable design, the frequency can be varied over a range as great as 10 to 1 by adjustment of the bias voltage.

Fig. 28C shows a cathode-coupled multivibrator circuit; in this circuit, RC coupling is provided from the plate of V_1 to the grid of V_2, exactly as in the plate-coupled multivibrator circuit in Fig. 28A. However, in Fig. 28C a cathode resistor (common to both tubes) has been substituted for the coupling capacitor between the plate of V_2 and the grid of V_1 in Fig. 28A. The frequency stability of this circuit may also be improved by returning the grids to a positive bias voltage. The waveforms and frequency of oscillation may be determined by the method of transient analysis previously described.

Synchronized Multivibrators. By use of a synchronizing voltage any multivibrator may be forced to oscillate above its "natural" frequency or at a submultiple of the synchronizing frequency. The output of a synchronized multivibrator is especially rich in harmonics and subharmonics, that is, contains frequencies that are multiples and submultiples of the synchronizing frequency. Harmonics as high as the 100th can be detected with the aid of tuned circuits without further amplification. By using a known frequency standard to stabilize the oscillations of a multivibrator, it is possible to maintain many frequencies related to the standard. In this way, the multivibrator circuit functions as either a frequency multipler or a frequency divider. For this reason, the multivibrator is very useful as a frequency-measuring device.

The control voltage, or synchronizing signal, may be introduced into the multivibrator circuit by any one of five common methods shown in Fig. 32; however, only one method can be used at a time.

The synchronizing signal tends to force the multivibrator to oscillate above its natural frequency. However, as the circuit is made to function at higher and higher frequencies, a point is reached beyond which the multivibrator synchronizes to one-half the control frequency. Still higher frequencies provide synchroniza-

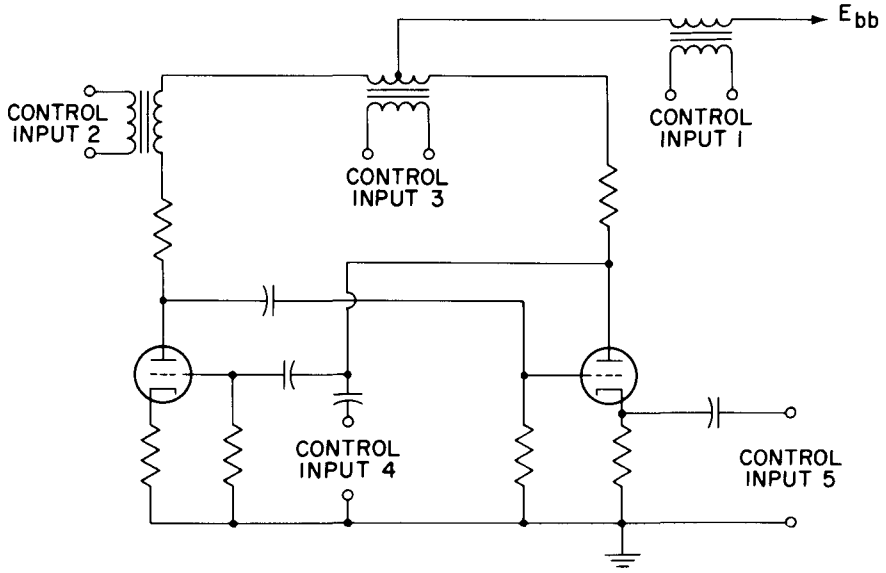


Figure 32. Five Methods of Synchronizing a Multivibrator

tion to one-fourth the control frequency, and so on. Similarly, the multivibrator may also synchronize to one-third, one-fifth, one-seventh, etc., of the control frequency.

If the multivibrator is balanced, the frequency at which it "locks" or synchronizes is 1/n times the control frequency (where n is any whole number). If the multivibrator is unbalanced, it locks at any multiple or submultiple of the control frequency. With perfect symmetry, a multivibrator may oscillate at either even or odd-frequency ratios, or show no discrimination, depending on whether the injected synchronizing signal is introduced in the same or opposite phases in the two tubes, or is introduced in only one tube. Common methods of synchronizing to submultiples of the control frequency are shown in Fig. 33.

Sawtooth Generators

Some of the more common generators of sawtooth waveforms are shown in Figs. 34 and 35. All circuits shown here make use of the changing potential across a capacitor which is being charged through a resistor from a constant potential source. The potential across the capacitor follows the form:

$$e_c = E_{bb}\,(1-e^{-t/RC}) \qquad (16)$$

where E_{bb} = supply voltage
t = time, in seconds
C = capacitance of capacitor, in farads
R = resistance of series resistor, in ohms

(A plot of this equation is shown in Fig. 36.)

As shown in Fig. 36, for a short time after the application of voltage, the change of voltage across the capacitor will be fairly linear with time. Thus, if the switching is properly timed, a sawtooth wave of reasonably good linearity may be obtained.

Gas-Tube Sawtooth Generators. One of the earliest and simplest sawtooth generators uses a glow-discharge or neon lamp, as shown in Fig. 34A. Assume there is no charge on the capacitor when the switch S is closed. The voltage across the capacitor will increase according to Eq. (16) until it reaches the breakdown potential of the gas in the neon lamp. The neon lamp will then conduct heavily, rapidly discharging the capacitor to the extinction potential of the neon lamp. The discharge will then cease and the charge across the capacitor will again rise according to Eq. (16) until the process repeats. The return time or discharge time of the sawtooth wave is very small, because of the high discharge current through the neon lamp. The frequency of oscillation is given by:

$$f = \frac{1}{RC \log_e \dfrac{E_{bb} - E_e}{E_{bb} - E_d}} \qquad (17)$$

where E_e is the extinction potential of the neon lamp and E_d is its breakdown potential.

The neon-lamp oscillator has at least three obvious shortcomings: (1) relatively high voltages must be used to obtain linear sawtooth waves; (2) the amplitude of the output wave is relatively small and not controllable; (3) the frequency of oscillation is somewhat unstable and not easily synchronized. These shortcomings may be largely overcome by substituting a thyratron for the neon lamp, as shown in Fig. 34B. In this circuit, R_g is a protective resistor to limit current in the grid circuit, R_p is a protective resistor to limit the peak plate current and R_L is a charging resistor for the capacitor C. If no synchronizing signal is applied to the control grid, the frequency of oscillation may be determined by Eq. (17) (in which R_L is substituted for R). However, with the thyratron it is possible to vary the breakdown potential E_d of the tube by varying the grid bias voltage E_{cc}. Thus, the frequency may be varied by variation of supply voltage E_{bb}, grid bias voltage E_{cc}, series resistance R_L, or shunt capacitance C.

If no synchronizing signal is applied to the control

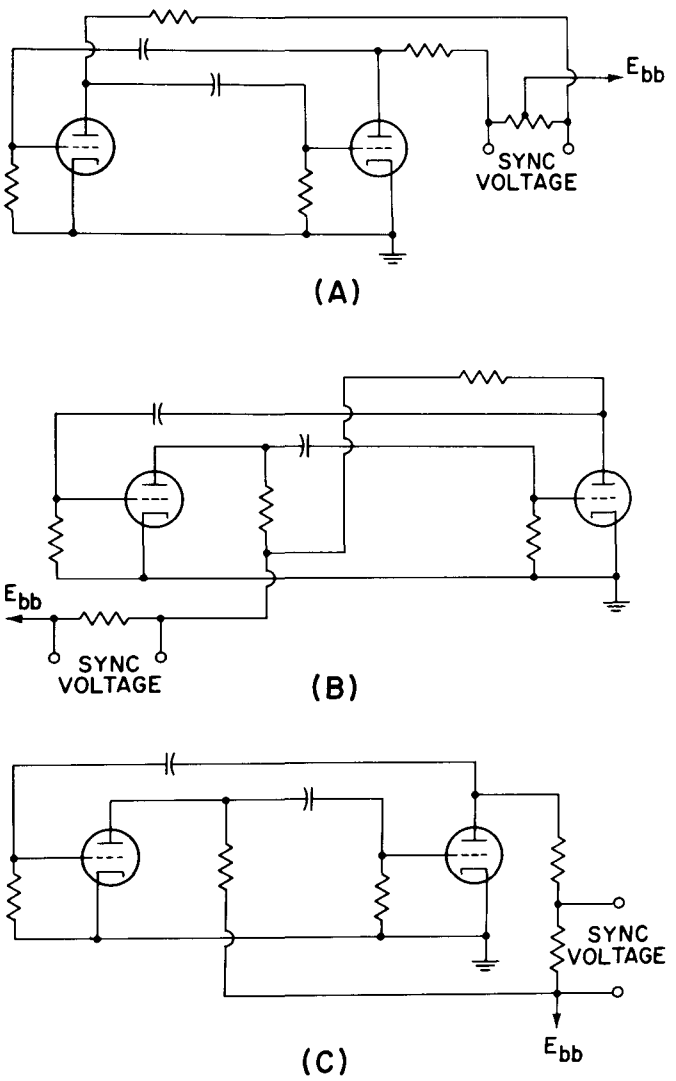


Figure 33. Circuits for Synchronizing to Submultiples of the Control Frequency; (A) Circuit favoring odd value of n; (B) Circuit favoring even value of n; (C) Circuit favoring any value of n

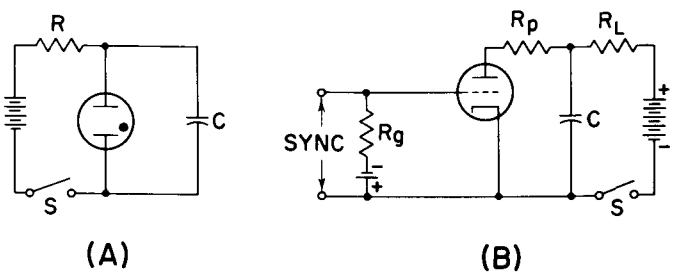


Figure 34. Gas-Tube Sawtooth Generators; (A) Neon-lamp oscillator, (B) Thyratron oscillator

grid, the thyratron sawtooth oscillator is no more stable than the neon-lamp oscillator. However, it may be synchronized at a given frequency by applying a small voltage at the desired frequency to the control grid. The manner in which synchronization is achieved is illustrated by Fig. 37.

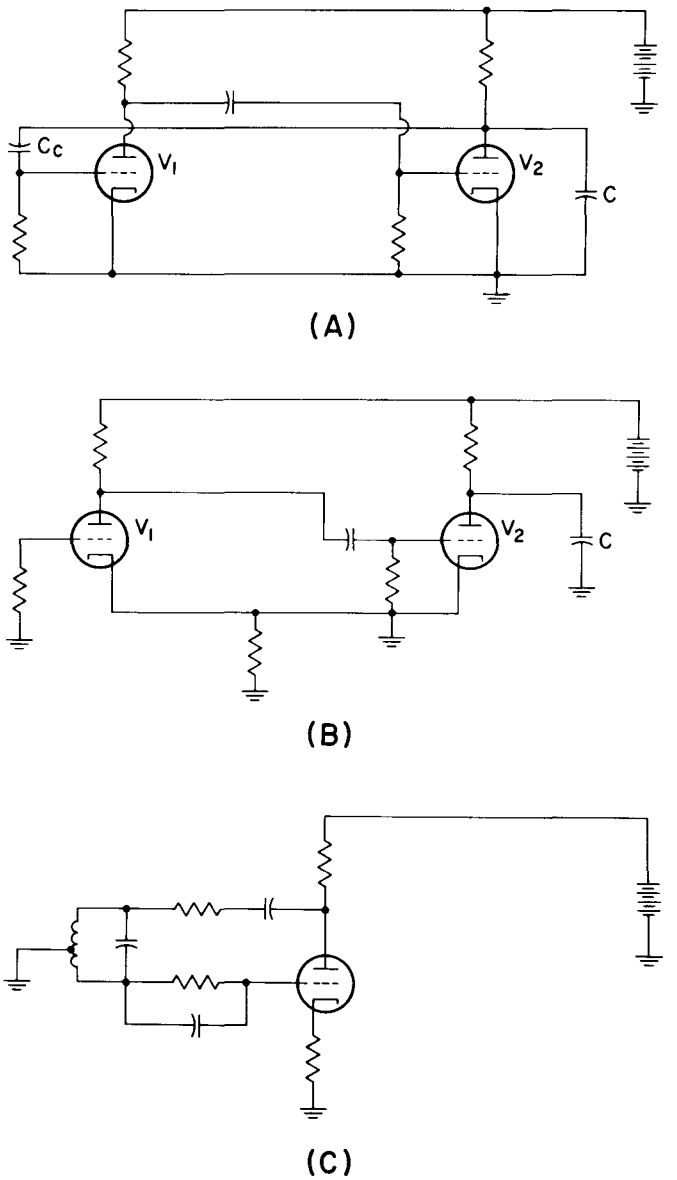


Figure 35. Free-Running Vacuum-Tube Sawtooth Generators: (A) Plate-coupled multivibrator type; (B) Cathode-coupled type; (C) Hartley oscillator type

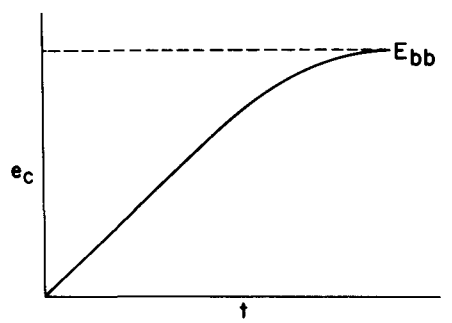


Figure 36. Changing Potential Across a Capacitor Being Charged Through a Resistor From a Constant Potential Source

The natural frequency of oscillation, as determined by Eq. (17), is adjusted to be slightly lower than that of

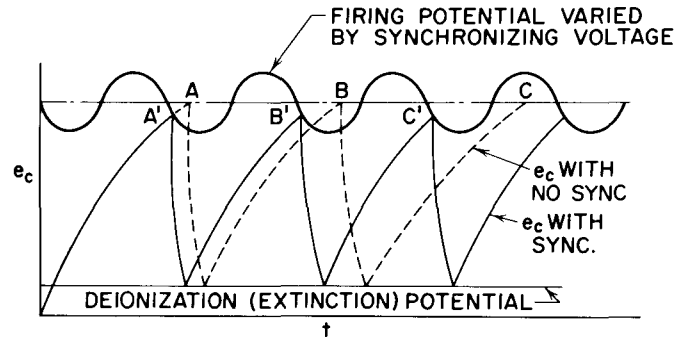


Figure 37. Effect of Synchronizing Signal on Frequency of Sawtooth-Wave

the synchronizing voltage. Without synchronization the tube would fire at points A, B, C, ... When a synchronizing signal is present, the firing point will vary in accordance with the grid signal to A', B', C', ... In a similar manner, the frequency may be synchronized to a multiple or submultiple of the desired frequency.

Vacuum-Tube Sawtooth Generators. The plate-coupled multivibrator type of sawtooth generator operates in the following manner. (See Fig. 35A) At the instant when the supply voltage E_{bb} is connected, the capacitor C will prevent any voltage from appearing at the plate of V_2. Therefore, V_1 will conduct heavily and V_2 will be cut off. The potential across capacitor C will rise at an exponential rate in accordance with Eq. (16) in exactly the same manner as with the neon-lamp sawtooth generator. At a certain potential across C, V_2 will start to conduct. The current through V_2 will be supplied by the charge built up in the capacitor C. The voltage across C will then decrease. This decreasing voltage across C (plate of V_2) will be coupled to the control grid of V_1 by means of the coupling capacitor C_c. The negative-going voltage at the grid of V_1 will result in decreased current through V_1 and a resultant increase in the voltage at the plate of V_1. This increase in voltage at the plate of V_1 is coupled to the grid of V_2 and results in an increase in the current through V_2. This action is cumulative and, therefore, capacitor C is discharged very rapidly through V_2. When C has discharged to a value at which V_2 will no longer conduct, V_1 again conducts and the process repeats.

The operation of the circuit of Fig. 35B is similar to that of Fig. 35A. However, while the coupling from the plate of V_1 to the grid of V_2 is capacitive, exactly as in Fig. 35A, the coupling from the plate of V_2 to the grid of V_1 is omitted. In its place, a cathode resistor, common to both tubes, is used. V_2 then operates as a cathode-follower and changes in its grid voltage also appear at its cathode and affect the grid-cathode voltage of V_1.

Synchronization of the frequency of oscillation of the circuits in Fig. 35A and 35B may be obtained by injecting a synchronizing voltage into either the cathode or grid circuits in a manner similar to that described for multivibrators.

The circuit in Fig. 35C illustrates one way of obtaining a sawtooth wave from a conventional LC oscillator circuit. This circuit is popular for electronic organs in which a waveform rich in harmonics is desired for achieving the various "voices" and in which the frequency stability of the LC oscillator is required. Actually, three waveforms are available from this circuit. A sine wave is available at the plate end of the oscillator coil, a pulse output is available across the resistor in the cathode circuit, and a sawtooth is available at the plate of the tube.

TUBES FOR USE IN OSCILLATOR CIRCUITS

High transconductance is required in any tube to be used as an oscillator. At very high frequencies where circuit losses are high and tube efficiencies low, transconductance must be higher than at the lower frequencies. In addition, the following factors are important in the choice (or design) of a tube for use as an oscillator: (1) available supply voltage, (2) power output required, (3) frequency of operation, and (4) type of oscillator (sinusoidal or non-sinusoidal voltage output).

Supply-Voltage Considerations

Supply voltage is important because it determines, to a degree, the amplification factor required in the tube and the cathode area required for a given power output. When complete freedom of design is given, the amplification factor of a tube is of very little importance in determining how well it will function as an oscillator. However, such freedom seldom exists in practice. The plate supply voltage, which is the most important factor, usually is fixed by other considerations. In general, if the supply voltage is low (as in portable or in radiosonde equipment) the amplification factor should be fairly low. Otherwise, insufficient plate current will flow to produce adequate power output. In addition, if the supply voltage is low it may be necessary to use a tube having a large cathode area and requiring high heater power in order to have sufficient cathode current available to produce the desired output. Conversely, if the supply voltage is high, the amplification factor should be high to prevent both excessive plate current and plate dissipation.

Power Output Considerations

The tube chosen for use as a power oscillator must have adequate cathode area and plate-dissipation capability to produce the required power output consistently, both initially and during life. The power requirements for local-oscillator use in radios and TV receivers are usually so small that there is seldom any problem in producing the required output. However, at ultra-high frequencies, the oscillator efficiency is so low that most of the plate input power is dissipated at the plate. In this situation, a plate-dissipation capability of two to three watts may be required for reliable operation, even though the power output required is only a few milliwatts.

Frequency Considerations

Vacuum tubes are normally required to operate as oscillators at frequencies from a few cycles up to about 1000 megacycles per second. In microwave applications (not considered here), the upper frequency may be many thousand megacycles per second. A tube which per-

forms very satisfactorily at low frequencies may be completely unusable at the high frequencies required of a local oscillator for a television receiver.

In general, all oscillators operate with peak plate currents considerably in excess of the average current. Thus, all tubes designed for oscillator operation must be capable of delivering high peak currents during the conducting portion of the cycle. This property is especially important at the higher frequencies, such as the uhf television band, because many factors, such as radiation losses, and dielectric losses in tube sockets and components tend to reduce the power output and efficiency of the oscillator. Any additional limitation on the capability of the oscillator, such as too low a peak current capability, tends to make oscillator operation unreliable. This problem tends to force the tube designer to use excessive cathode temperature in tubes designed for uhf operation. The high cathode temperature, however, may produce an effect exactly opposite to that desired. Although the peak current capability may be very high, the high cathode temperature can cause sublimation of cathode base metal on the micas and stem of the tube where it produces leakage paths. These leakage paths appear as a low resistance in shunt with the tuned circuit and cause severe loading; in some cases, the effect is to stop oscillation completely. The material deposited on the micas can also change tube capacitances radically and cause a shift in the frequency of oscillation. For longest life, it is better to maintain normal cathode temperature and adjust the tube processing to obtain the required peak current capability.

Tube capacitances and lead inductances are of little importance at low frequencies. At ultra-high frequencies, however, they may become major frequency-determining elements. For example, the internal lead inductances and capacitances of type 6AF4-A (together with those of the associated tube socket) are such that the self-resonant frequency of the tube is about 1200 megacycles when a short is connected between plate and grid. This high resonant frequency is achieved by the use of multiple connections to the plate and grid electrodes, low mount height to reduce lead inductances, and small electrodes to reduce capacitances. While it is possible to obtain oscillations above the self-resonant frequency of the tube by series tuning, this method is usually not practical. Most practical oscillator circuits are designed to operate somewhat below the self-resonant frequency of the tube. A further increase in the self-resonant frequency of the tubes may be obtained by the use of special tube constructions such as that used in pencil tubes and lighthouse tubes in which the electrodes are attached to heavy disks and cylinders instead of small stem leads.

Input loading is another factor which, although of no concern at low frequencies, is quite important at ultra-high frequencies. The loading produced by electron-transit time and by cathode-lead inductance appears as a shunt resistance across the input circuit of the tube and reduces the oscillator output and efficiency. Therefore, for best operation at high frequencies, close interelectrode spacings and fine grid wires are necessary. In addition, for best circuit-loading characteristics, the grid should be shaped to match the contour of the cathode.

Sinusoidal and Nonsinusoidal Oscillator Considerations

The requirements for tubes for sinusoidal output have been fairly well covered in the preceding discussion. Cutoff is of some importance because a tube with poor cutoff is likely to have poor input loading and somewhat poorer frequency stability. Cutoff is extremely important in the case of nonsinusoidal oscillators such as multivibrators. If the cutoff is sharp and well defined, the frequency stability of the multivibrator will be much better than if the cutoff is remote and poorly defined.

It is common for multivibrators and other relaxation oscillators to use extremely high circuit impedances; plate and grid resistors of 0.5 to 10 megohms are normal. With such high circuit impedance, it is important that interelement leakage resistances be very high initially and during life. Leakage resistances should be several hundred megohms initially and remain above 100 megohms during life. This requirement dictates the use of cathode material which does not vaporize easily as well as a cathode temperature which is no higher than necessary for satisfactory operation. In addition, if feasible, leakage slots should be incorporated in the micas and the micas should be sprayed with alundum or other insulating material to lengthen the leakage paths.

In many applications of multivibrators, it is desirable to make the rise time of the output waveform as short as possible (a fraction of a microsecond). If the input capacitance of the tube used is high, the grid voltage cannot change rapidly and, therefore, the rise time will be lengthened. Thus, it is desirable to use tubes with small interelectrode capacitances for such applications. In addition, the use of tubes with high transconductance will aid in obtaining a rapid transfer of conduction from one tube to the other.

REFERENCES

1. Grammer, G., "The Dynatron Frequency Meter," QST, Vol. 14, p. 9, Oct. 1930
2. Iinuma, H., "A Method of Measuring the Radio-Frequency Resistance of an Oscillatory Circuit," Proc. IRE, Vol. 18, p. 537, Mar. 1930
3. "The Use of the 57 or 6C6 to Obtain Negative Transconductance and Negative Resistance," RCA Application Note No. 45
4. Reich, Herbert J., "A Low Distortion Audio-Frequency Oscillator," Proc. IRE , Volume 25, p. 1387, Nov. 1937
5. Pierce, G. W., "Magnetostriction Oscillators," Proc. IRE, Vol. 17, p. 42, Jan. 1929
6. Owens, J. H., "Audio Oscillator Uses New RC Design," Electronics, March 1954
7. Code 01
8. "Klystrons and Microwave Triodes," Vol. 7, MIT Radiation Laboratory Series, p. 179, McGraw-Hill

Beam Power Tube Design Considerations

O. H. Schade, Jr.

Harrison

This article, which is concerned with the design of beam power tubes, begins with a section on space-charge effects and their influence on the plate-current characteristics of the beam tube. The next section is concerned with the electron beam-characteristics at zero control-grid potential. Both sections relate the important parameters which affect the amount of screen-grid current drawn by a beam power tube. Methods for calculating the approximate beam shape at zero grid potential are then presented together with an examination of the agreement between measurement and calculation. The next section covers the variation of screen-grid current with control-grid potential and explains the effects of beam formation at other than zero control-grid potential. High-voltage phenomena and considerations are discussed for those tubes which operate with large plate-voltage pulses. The article concludes with a discussion of a combined mechanical-thermal-electrical design procedure suitable for beam power tubes. Specific references mentioned in the text are compiled at the end of the article, along with others of general value.

A beam power tube can generally be operated more efficiently in power output service than a pentode for two main reasons: (1) relatively high plate currents can be obtained at plate potentials considerably below the screen grid potential; (2) the screen grid of a beam power tube collects a relatively small portion of the cathode current. The current flowing between the screen grid and plate is sufficient to cause a large depression in the space potential between them. When this potential minimum occurs uniformly throughout the area of current flow, is large enough in magnitude, and is properly positioned with respect to the plate, the majority of the secondary electrons from the plate surface cannot penetrate this potential barrier, and hence, are returned to the plate. A pentode begins to lose plate current as soon as the plate potential falls below that of the screen grid.

The screen grid collects a relatively small portion of the cathode current because the screen grid is aligned with the control grid. When the screen-grid and control-grid wires are aligned and properly positioned, it is possible to cause a cathode current flow in directed beams having the uniform current density necessary for the formation of a uniform potential minimum. The beams also serve to direct the majority of the cathode current between the screen-grid wires. The pentode structure also forms current beams, but electrons approach the screen-grid plane in a more random fashion (due to the absence of grid-wire alignment) and, therefore, a considerable number are pulled directly to the screen-grid wires. In addition, a large number of the electrons which pass through the screen-grid plane approach the suppressor grid with insufficient forward velocity-components to pass through the suppressor-grid plane and reach the plate.

The design of a good beam power tube requires an understanding of beam formation and the space-charge effects in the region beyond the screen grid. Beam formation is usually analyzed with the aid of field plots. Space-charge effects are usually so complex that, if an idealized plane-parallel treatment of the problem does not yield sufficiently accurate results, an orderly "trial-and-error design" of developmental structures is most practical. The work in this article neglects the effect of initial velocities of the emitted electrons, and the effects of space charge between control grid and cathode upon the electron trajectories, and is, therefore, insufficient to explain some of the phenomena associated with tube operation at very low voltages. The following three sections are concerned with space-charge effects and beam characteristics.

THE BEAM POWER TUBE IN CROSS SECTION

Fig. 1 shows the field configuration and electron paths in a typical beam power tube. The equipotential lines were determined from the resistance-paper field-plot, and electron trajectories were traced on this plot. Initial velocities were neglected and the space-charge minimum position was approximated by inserting a zero-potential plane at the beam-confining plate position. Note that the electrons leaving the cathode converge due to the field shape, that the beam cross-section is a minimum near the screen-grid plane. Those electrons having large velocity components tangential to the cathode do not penetrate the potential minimum, but return eventually to the screen-grid wires. Also, the field at the cathode is not uniform, so that the current density in the outer electron rays is lower than that in the rays near the center.

The two general factors of prime importance are the characteristics of the beam, and the effect of the potential minimum upon this beam. The space-charge effects will be treated first.

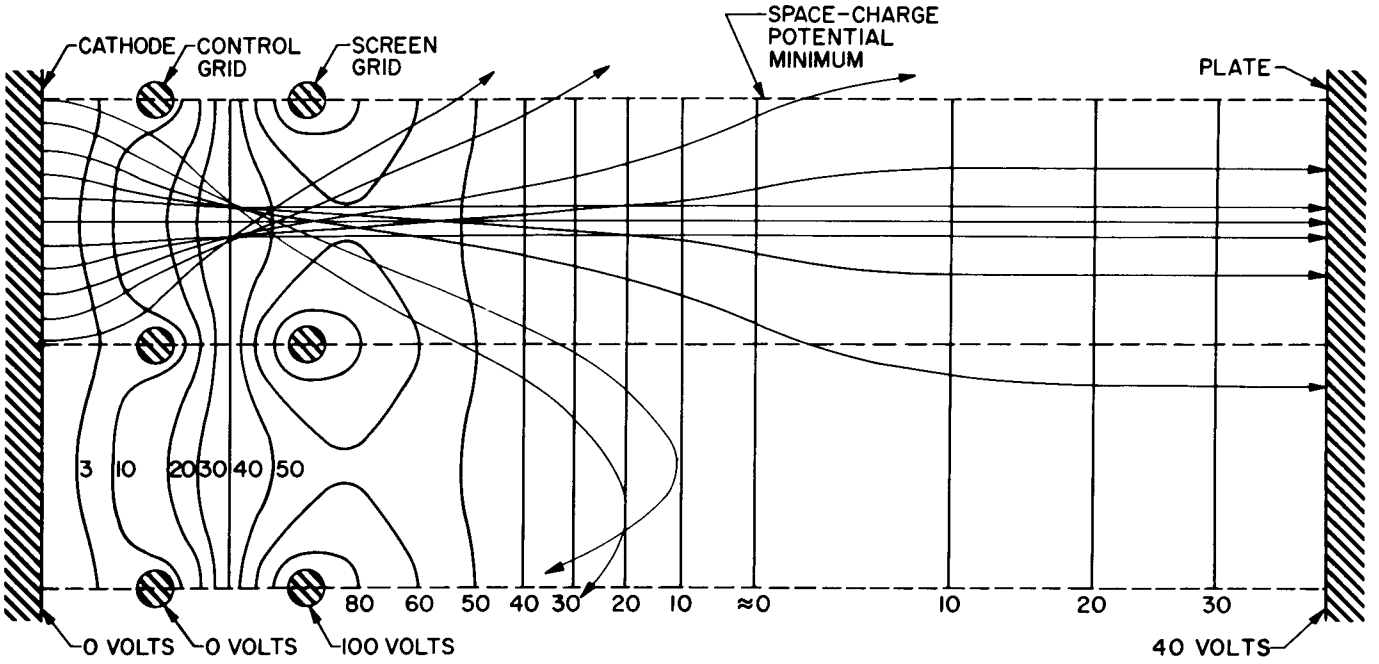


Figure 1. Field Shape and Electron Trajectories in a Typical Tube Operated with the Control Grid at Cathode Potential and the Plate Potential Lower than the Screen-Grid Potential

SPACE-CHARGE EFFECTS[1, 2, 3]

The potential distribution in the screen-grid-plate region depends upon several factors, some of which are: (1) The screen-grid and plate potentials, and the separation of these electrodes; (2) The injected current density; and (3) The amount of the current injected at the screen-grid plane that gets through the potential minimum and reaches the plate.

The ideal plane-parallel case, with normal injection and uniform velocity at the screen grid plane will be treated. Under these conditions, the potential distributions of interest are shown in Figs. 2 and 3. Their source and the nomenclature used are from Spangenberg,[2] who has classified three main types of potential distribution which may exist in the screen-grid to-plate region:

1. Type A distribution, which is concerned with the potential distribution formed with negative plate potentials.
2. Type B distribution, which exists when the plate potential lies between zero volts and that potential at which all the injected current is transmitted to the plate.
3. Type C distribution, which exists for all plate potentials above that potential which is necessary for full transmission of the injected current. The parameter τ in Fig. 2 is the percentage of the injected current which is transmitted through the potential minimum. The designation ϕ_{min} in Fig. 3 refers to the normalized magnitude of the potential minimum for that particular curve.

The ordinate in both figures is the ratio of the potential of the plate to that in the screen-grid plane. The abscissa

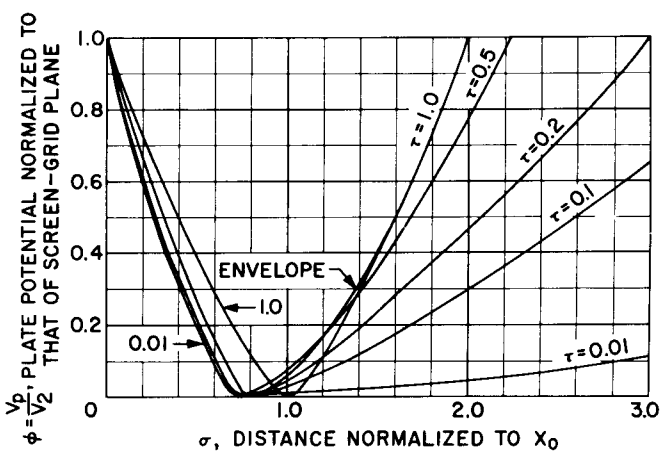


Figure 2. Potential Distribution Curves, Type B

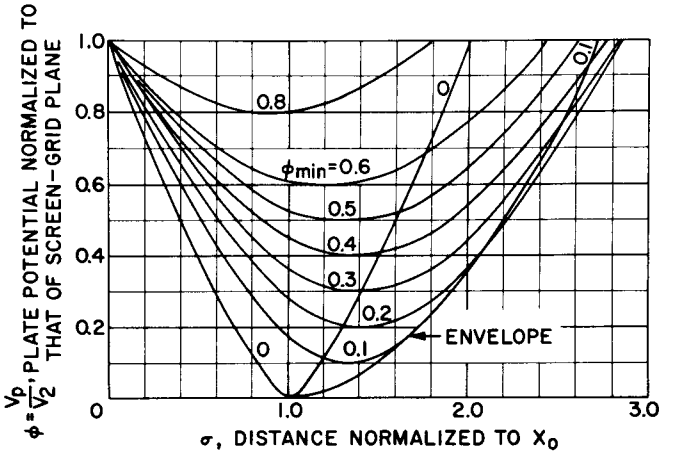


Figure 3. Potential Distribution Curves, Type C

in both figures is distance normalized to a distance X_0, the distance between screen-grid plane and the zero-potential virtual cathode when 100 per cent current transmission through the minimum exists. If the virtual cathode and screen-grid are considered as a diode, X_0 may be found from the relationship

$$X_0 = \sqrt{\frac{2.33 \times 10^{-6}}{J}} V_2^{3/4} \text{ cm} \qquad (1)$$

where J is the current density in amperes per square centimeter, and V_2 is the potential in volts in the screen-grid plane. For a particular current density and potential, the screen-grid-plate spacing is expressed in units of σ. If, for example, $X_0 = 0.1$ cm for a particular calculation, and the screen-grid-plate spacing is 0.14 cm. then $\sigma = 1.4$. The potential distributions shown in Figs. 2 and 3 apply for the case where $0 \leqq E_b \leqq V_2$, where E_b is measured along the line 1.4σ for this example.

Fig. 4A shows some of the potential distributions when the plate potential is increased from 12 to 87 volts (V_2 = 100 volts). The sequence of curves a, b, c is followed until the plate is at 30 volts potential. At this point, the current transmission reaches 100 per cent, and the virtual cathode can no longer be sustained at zero potential. The distribution therefore jumps to type C as shown by the curve c', and remains type C through curves d and e. When the plate potential is lowered, the distributions shown in Fig. 4B are encountered. The type C distribution is maintained until the envelope of Fig. 3 is reached at curve h. At this point, the distribution changes back to type B, curve h'.

Fig. 5 shows the current transmission as a function of plate potential for the structure of Fig. 4. It can be seen that the knee of the curve occurs at 30 volts when the plate potential is rising. However, 100 per cent current transmission is maintained down to eight volts once a type C distribution is established. The resultant characteristic should be a familiar one to the beam tube designer and user, even though other factors alter the shape somewhat and the "shift" of the knee voltage is in most cases much smaller in actual beam tubes.

One of the factors which causes a departure from the "ideal" characteristics is secondary emission from the plate.[4] Further examination of the potential distributions of Fig. 4A and 4B points out that, once the plate potential is above 30 volts, this structure does not permit the formation of a space-charge minimum much lower in potential than the plate potential. Even when the plate is at 87 volts, the minimum is 80 volts. The difference of seven volts is probably just sufficient to effectively suppress secondary electrons (figures of 10 to 20 volts are found in the literature). Secondary emission will put a characteristic "undersuppressed" trough in the region from c to e of the Fig. 5 curve.

Secondary-emission effects could be reduced by increasing the screen-grid-to-plate from 1.4σ to 1.8σ. The distributions would then be as shown in Fig. 4C. Notice that the minimum is always at least seven volts lower than the plate potential. However, the change from a type B to a type C distribution does not occur

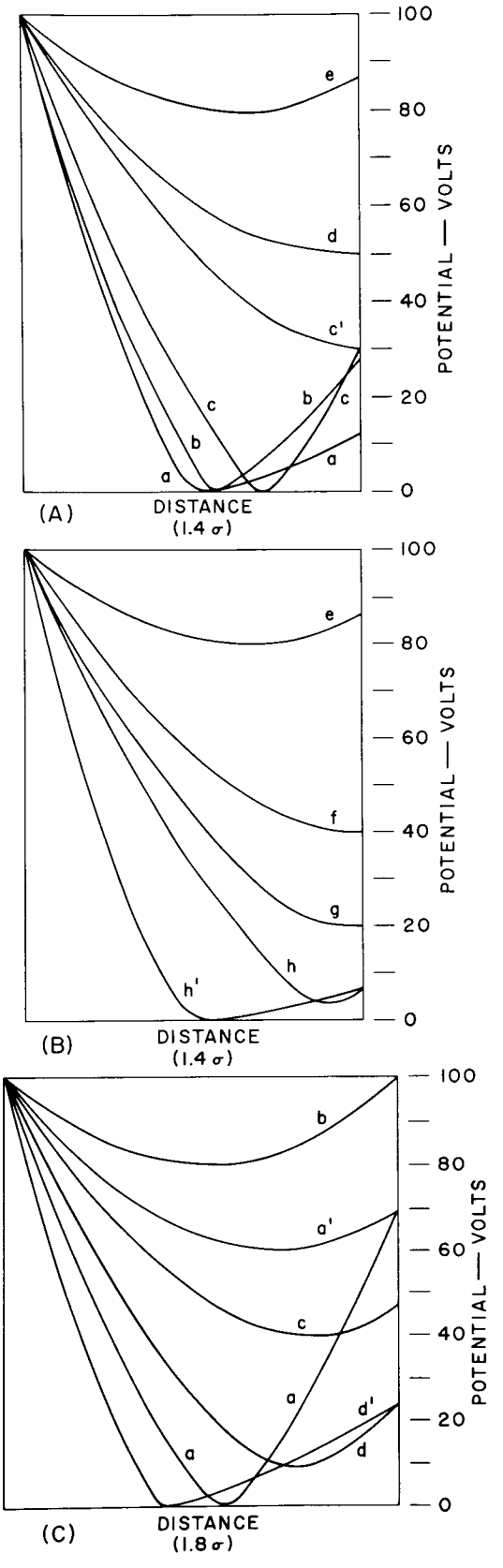


Figure 4. Potential Distributions for Constant Injected Current: (A) sequence of increasing plate voltages — 1.4σ; (B) sequence of decreasing plate voltages — 1.4σ; (C) distributions with spacing of 1.8σ

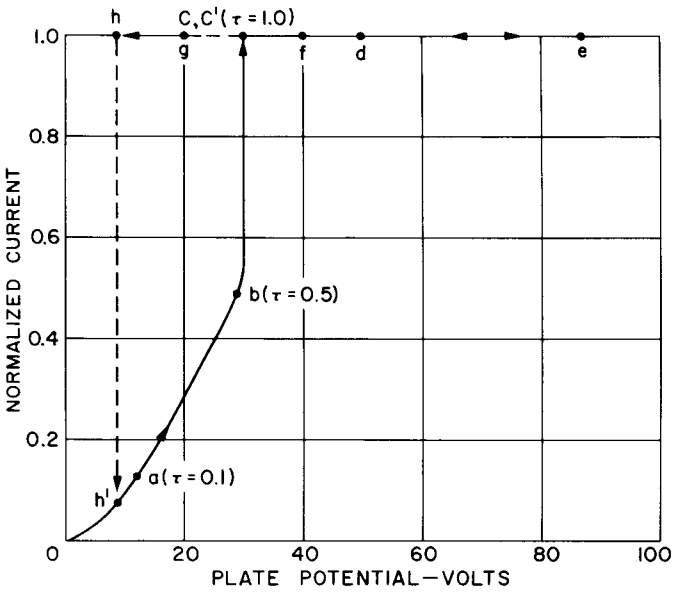


Figure 5. Plate Current as a Function of Plate Voltage in the Ideal, Plane-Parallel Beam Tube

until the plate potential is 70 volts. Freedom from secondary-emission effects has been assured only by accepting a relatively high knee voltage. Although it may not be necessary to make the spacing as large as 1.8 σ in the practical tube, the theory illustrates that too small a screen-grid-to-plate spacing leads to secondary-emission effects in the tube characteristic.

Fig. 6 shows the changes made in the plate-current characteristic of a developmental tube when the screen-grid-to-plate spacing is changed. Curves a, b, and c follow the sequence of decreased spacing. The curves illustrate that when the screen-grid-to-plate spacing is the only parameter varied, lower knee voltages are accompanied by a drop in plate current and plate resistance. If the spacing is too small, as in curve c, secondary-emission effects are observed. Notice also that the screen-grid current increases as the plate voltage

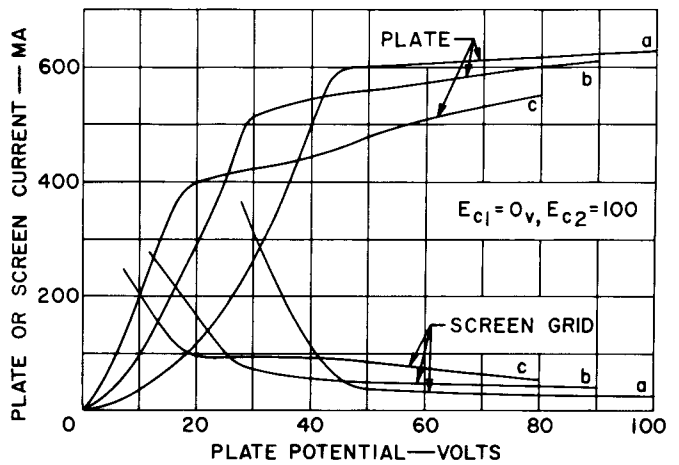


Figure 6. Plate and Screen-Grid Currents as a Function of Plate Potential, with Different Plate Positions for a Developmental Tube

is dropped. Part of the increase may be due to secondary-emission, as may be seen in curve c, but the greatest part is generally due to the effects of the decelerating field between screen grid and virtual cathode. This point will be covered later in the article.

Table I gives the approximate values of the parameters which affect the knee voltage E_{b0}. The injected current has been approximated at 600 milliamperes for the three versions. When this quantity is divided by the virtual-cathode area (the beam confining plate "window" area was used), the resulting current density is found to be 85 milliamperes per square centimeter. The average potential in the screen-grid plane is in the order of 70 volts when the wires are at 100 volts and the plate potential is about 50 volts. It will be assumed that V_2 in Eq. (1) is a constant 70 volts for all three versions. The value of X_0 is found to be about 50 mils. The screen-grid-to-plate spacings in the three versions yield the values of σ found in the table, from which the envelope of Fig. 2 may be referred to for the values of E_{b0}. The agreement between observed and calculated knee voltage is reasonably good—better than can be expected in the usual case. The low σ-value of 1.35 for curve c indicates that secondary emission should influence the characteristic. This effect can be seen on the characteristic. There is, however, no appreciable effect at 1.55 σ.

Table I

Approximate Values of the Parameters Which Affect the Knee Voltage E_{b0} of a Developmental Tube

Curve	Injected Current		X_0 mils	σ	ρ	E_{b0} (calc) volts
	I ma.	J ma./cm²				
a	600	85	50	1.75	3.15	49
b	600	85	50	1.55	2.8	32
c	600	85	50	1.35	2.45	20

The quantity ρ is also shown in Table I. This quantity is the ratio of the screen-grid-to-plate spacing to the screen-grid-to-cathode spacing[1]. Most beam power tubes made for output service have values of ρ lying between 2.9 and 4.1. The lower values will be found in tubes where the operating screen-grid potential is relatively high (250 volts or more), and also in the latest low triode mu types of horizontal output tubes. Medium to large values of ρ will be found in many tubes where screen grid dissipations dictated the need for better plate-to-screen-grid current ratios. In general, tubes with low values of ρ must be more carefully designed, because the knee region is more sensitive to the uniformity of beam-current density, dimensions of the beam confining plates, "flatness" or "taper" of the active plate surface, and the beam shape.

The position of the virtual cathode changes when the injected-current density changes. A decreasing in-

jected-current density causes the virtual cathode to move toward the plate and thereby, lowers the knee of the plate-current characteristic. This effect leads to the characteristics shown in Fig. 7, which was obtained on a developmental space-charge tetrode designed by H. Ruzinsky of the Entertainment Tube Design Section. This tube has no beam confining plates and is a "true" space-charge tetrode. The plate cross-section is U-shaped on each side of the cathode, with a concave "dent" in the bottom of each U to keep secondary electrons within the beam so that they will be returned to the plate. Because the structure lacks beam-confining plates, the potential distribution between screen grid and plate is the closest to the theoretical condition that has yet been found in a practical tube design. The degree to which the knees are displaced is, therefore, about the largest which could be expected in a practical design. Many tubes show much smaller displacements of the diode lines. Table II gives the approximate values of the parameters which affect the knee voltage at different values of control-grid potential. The calculated knee voltages have a larger spread than the observed values. There are several reasons for this:

1. The system is not plane-parallel, nor are all the electrons injected normal to the screen-grid plane, nor is the screen-grid-plane potential uniform or constant.
2. When the control-grid potential is decreased, the beam width is reduced due to the effects on the field of the control-grid siderods. This action tends to maintain the beam-current density, and therefore, there is less movement of the space-charge minimum than has been calculated.
3. Because the active plate surface is concave, an average distance for the screen-grid-to-plate spacing was chosen. When the virtual cathode approaches the plate, the diode formed has a very nonuniform spacing, a condition which has not been accounted for in the calculations. This condition is of particular importance, for example, at the control-grid potential of -20 volts where the average value of σ is 1.03, and the potential minimum does not exist in some portions of the beam.

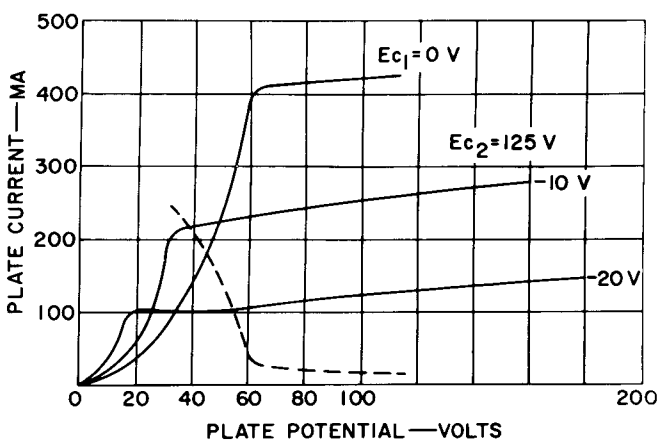


Figure 7. Plate-Current Characteristics of a Developmental Space-Charge-Suppressed Tetrode Having no Beam-Confining Plates

Table II

Approximate Values of the Parameters Which Affect the Knee at Different Values of Control-Grid Voltage

Control-Grid Voltage	Injected Current		X_0 mils	σ	E_b (calc.) volts
	I ma.	J ma./cm^2			
0	400	108	51	2.05	85
-10	220	60	69	1.50	34
-20	100	27	102	1.03	7

The largest spread in the knee voltage at different control-grid potentials will be found in tubes which have uniform beam-current densities and proper position of the edges of the beam-confining plates. The effect is particularly noticeable when the window of the beam-confining plates is relatively wide.

The dimensions of the beam-confining plates should be such that they form a continuation of the virtual cathode beyond the beam, so that secondary electrons cannot bypass the potential barrier and strike the screen grid. The beam-confining plate-screen-grid spacing is determined by the virtual-cathode position which corresponds to the beam-current density at which the tube will be operated. In many cases, this position places the window of the beam-confining plates roughly halfway between screen grid and plate, perhaps a little closer to the plate. The window opening should be as wide as the beam which would be formed without beam-confining plates. This width is usually at least as large as the center-to-center dimension of the control-grid siderods. The numerous deviations from the theoretical case in a practical tube make the design of beam-confining plates a somewhat trial-and-error affair.

THE ELECTRON-BEAM CHARACTERISTICS

Electron beams of uniform density, injected normal to the screen-grid plane, do not occur in beam power tubes, because the cathode, control grid, and screen grid form an electron-optical system which (with proper applied potentials) causes the electrons leaving the cathode to converge to line foci as indicated in Fig. 8. This figure shows the trajectories of electrons emitted normal to the cathode surface in a developmental tube that is operated with cathode and control grid at zero potential. Because the gradient is not constant over the cathode surface, the current density is not constant through any beam cross-section examined, and injection at the screen-grid plane varies considerably from the normal.

The converging beam does have an advantage, however, because the electron-optics may be chosen so that the screen-grid wires intercept very little current directly from the cathode. For example, Fig. 8 indicates that none of the electrons emitted normal to the

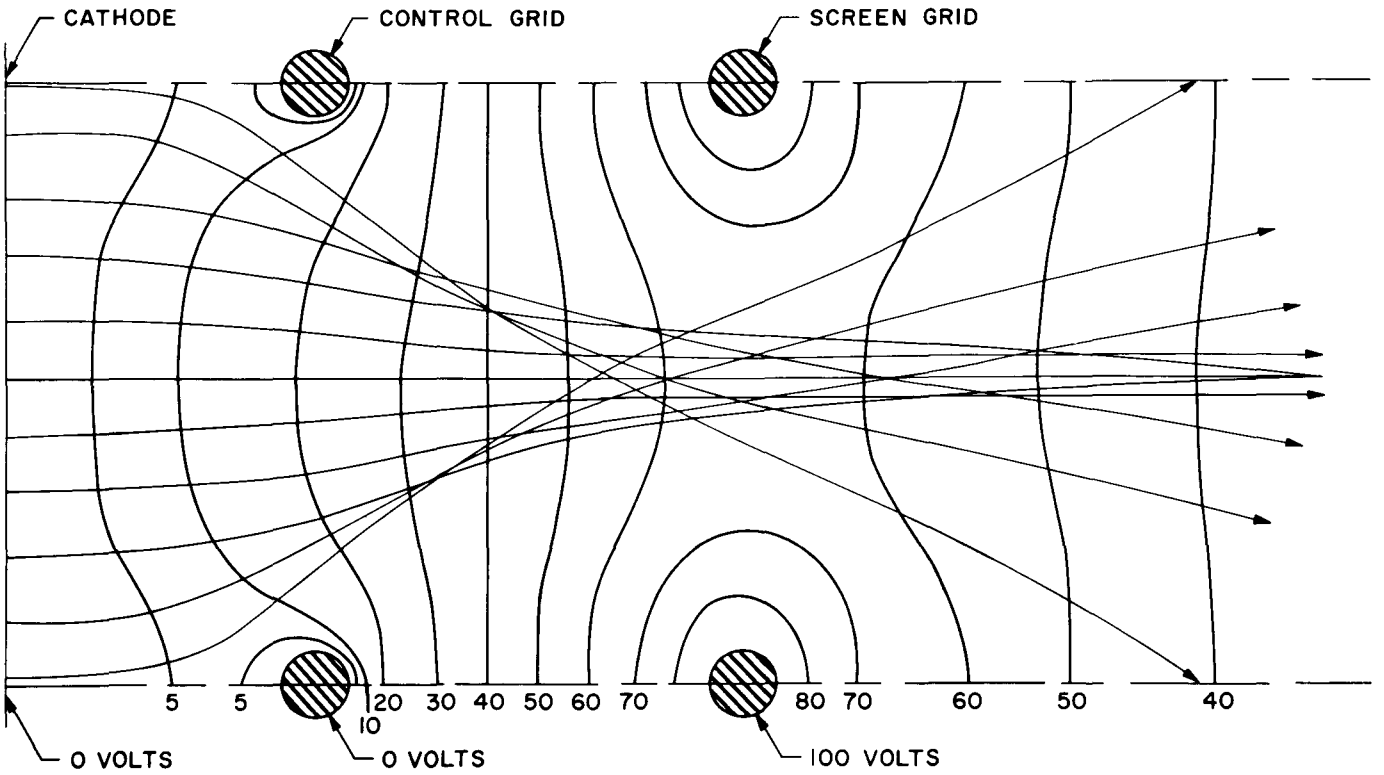


Figure 8. Electron Trajectories for a Developmental Tube

cathode surface of this structure are intercepted by the screen-grid wires. The beam cross section is at a minimum very near the screen-grid plane at zero control-grid potential. The measured current directly intercepted by the screen-grid wires of this structure is about 3 per cent of the total cathode current, and is due to electrons which leave the cathode in directions other than normal to its surface.

The ideal beam formation would be one which has just sufficient convergence to assure a minimum of current interception by the screen-grid wires, and then injects current at the screen-grid plane with just sufficient divergence to form a continuous space-charge minimum at the virtual cathode. This condition may be approached when the focal length of the beam is large compared to the pitch of the grids, or when the field shape is such that no real line focus is ever formed. Such a structure would require grids having a relatively small pitch and extremely fine wires. Because of the limitations inherent in the mechanical and thermal properties of the grid structures, considerable development would be necessary to achieve the ideal beam formation.

When electrons enter the screen-grid plane at angles to the normal, they are no longer capable of penetrating a virtual cathode at zero potential (neglecting initial velocities). The greater the value of the potential minimum, the larger may be the angle of entry that will allow penetration. The approximate value of potential that will allow the penetration of electrons with angular velocities may be determined by considering the electron behavior in an isolated linear retarding field, as shown in Fig. 9. Here, electrons are injected with a volt-velocity V_2 (100 volts in the figure) into the isolated field bounded at the other end by the potential V_{min} (25 volts in the figure). The trajectories within

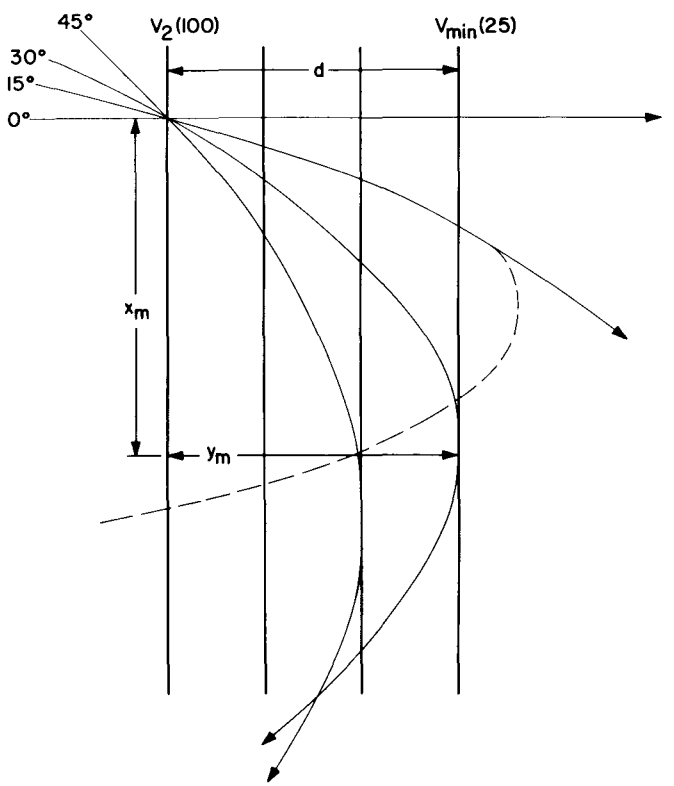


Figure 9. Electron Trajectories in Linear Retarding Field

the boundaries of the field are parabolic, and may be expressed*

$$y = \frac{-Ex^2}{4V_2 \sin \alpha} + \frac{x}{\tan \alpha} \qquad (2)$$

where α is the entrance angle and E is the field gradient, which has the value

$$-E = \frac{V_2 - V_{min}}{d} \qquad (3)$$

The coordinates of the origin of the parabola are:

$$x_m = \frac{2V_2}{E} \sin \alpha \cos \alpha \qquad (4)$$

$$y_m = \frac{V_2}{E} \cos^2 \alpha \qquad (5)$$

For the example shown in the figure, it can be seen that an electron entering the field at the limiting angle $\alpha_L = 30^o$ just reaches the boundary at V_{min}. A bit of manipulation produces the expression for the limiting angle in terms of the boundary potentials:

$$\cos^2 \alpha_L = \frac{V_2 - V_{min}}{V_2} \qquad (6)$$

This expression is plotted with V_2 as a parameter in Fig. 10. This figure permits the determination of the approximate value of the potential minimum which will allow 100 per cent current transmission through the minimum when the entrance angle and potential are specified. A minimum plate potential of 25 volts, for example, would theoretically allow 100 per cent transmission of a 30 degree beam at a screen-grid-plane potential of V_2 = 100 volts. Any electron entering at an angle smaller than 30 degrees would reach the plate, while those with larger angles would be returned to the screen grid. The practical tube will require an additional potential (of the order of 10 volts) on its plate to suppress secondary electrons.

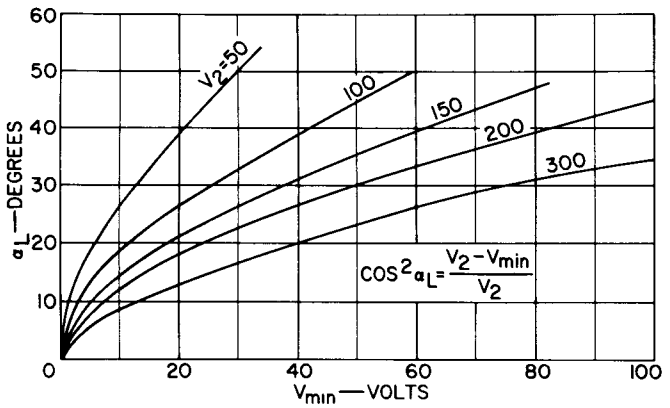
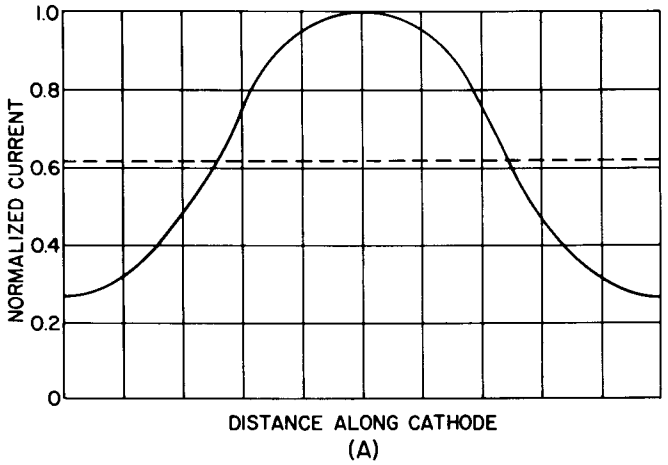


Figure 10. Limiting Entrance Angle for 100 Per Cent Transmission Through a Linear Retarding Field

*See Ref. 2, pg. 100.

If it is assumed that all of the cathode current is carried along the electron trajectories shown in Fig. 8, the current distribution as a function of entrance angle at the screen-grid plane can be determined as shown in Fig. 11B. It should be cautioned that the distribution shown is composed of current bundles having discrete angles, and is not continuous. Fig. 11B was constructed by determining the variation of mu over the cathode surface*, computing the currents in each ray, and normalizing the results as shown in Fig. 11A. It can be seen that if the current in the outer 24-degree ray were returned to the screen-grid wires (and 3 per cent for



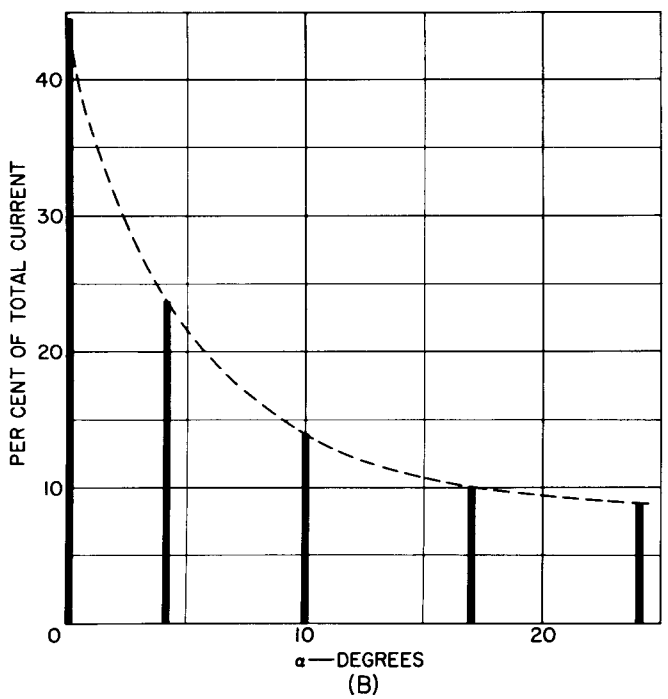


Figure 11. Relative Magnitude of Current in the Beam Rays: (A) distribution of current along the cathode surface; (B) entrance angles of current at screen-grid plane

*See article on "The Electric Field and Space Current."

direct interception were added), the screen-grid current would be in the order of that found by measurement, i.e., 6 to 11 per cent for many horizontal deflection tubes. To accurately predict screen-grid current by this method, more rays should be plotted. The direct interception should also be determined by tracing rays which leave the cathode at angles other than the normal.

CALCULATION OF THE APPROXIMATE BEAM SHAPE AT ZERO CONTROL-GRID POTENTIAL

It is possible to calculate the approximate beam shape by applying the laws of electron ballistics in electrostatic fields and an equation for the focal-length of a two-dimensional electron-optical system. The following analysis assumes that a "thin" electron-optical system exists; that is, that electron deflection occurs within a very small path length at the grid aperture. This is not true in a practical beam power tube, as can be seen from the traced electron trajectories. It is possible to treat the structure as a thick electron-optical system, but the author's attempts to simplify the problem to the point where a mathematical solution is less time-consuming than a trajectory-tracing have not resulted in a consistent approach (or result, for that matter)[7].

When a thin electron-optical lens is formed by a rectangular aperture positioned between uniform fields, rays parallel to the system axis tend to converge to an apparent line focus distant from the plane of the aperture as defined by the equation

$$f_a = \frac{2\,V_1}{E_2 - E_1} \tag{7}$$

which was proposed by Davisson and Calbick.[5] The quantity f_a is the apparent focal length of the lens; V_1 is the potential in the plane of the aperture; and E_1 and E_2 are, respectively, the field gradients in front of and behind the aperture. This equation does not take into account the effect of the accelerating field beyond the aperture, as was pointed out by Bull, who considered the electron-optical system formed in tubes with aligned grid.[6] Bull proposed an equation for the "corrected" focal length of the form

$$f_0 = f_a \left[1 + \frac{f_a}{4s} \left(\frac{V_2}{V_1} - 1 \right) \right] \tag{8}$$

This equation neglects aberrations, assumes that the lens is thin and that the field beyond the aperture is essentially uniform and free from space-charge effects. The quantity S is the control-grid-to-screen-grid spacing, and V_2 is the potential in the screen-grid plane. The potentials V_1 and V_2 may be removed from the expression for the potential in the control-grid plane

$$V_1 = \frac{E_{c1} + \frac{V_2}{\mu}}{1 + \frac{1}{\mu} + \frac{s}{\mu a}} \tag{9}$$

where a is the control-grid-to-cathode spacing. Eq. (9) may be reduced to the zero control-grid potential ratio

$$\frac{V_2}{V_1} = 1 + \mu + \frac{s}{a} \tag{10}$$

The corrected focal length may then be expressed

$$f_0 = f_a \left[1 + \frac{f_a}{4a} \left(\frac{\mu a}{s} + 1 \right) \right] \tag{11}$$

To remove the potential terms from equation (7), the gradients may be expressed in terms of V_1 and V_2;

$$E_1 = \frac{4}{3} \frac{V_1}{a} \tag{12}$$

$$E_2 = \frac{V_2 - V_1}{S} \tag{13}$$

When Eqs. (10), (12), and (13) are substituted into equation (7), the potential terms may be canceled out, and the expression becomes

$$f_a = \frac{2a}{\left(\frac{\mu a}{s} - \frac{1}{3} \right)} \tag{14}$$

The corrected focal length then becomes

$$f_0 = \frac{2a}{\left(\frac{\mu a}{s} - \frac{1}{3} \right)} \left[1 + \frac{1}{2} \frac{\left(\frac{\mu a}{s} + 1 \right)}{\left(\frac{\mu a}{s} - \frac{1}{3} \right)} \right] \tag{15}$$

It should be noted that the term $\mu a/s$ is the amplification factor from control-grid to cathode, which is commonly called the "reverse mu." Long focal-length beams will be obtained with low reverse mu and large control-grid-to-cathode spacings. The desirable small grid-to-cathode spacing needed for good perveance forces the use of relatively small control-grid-wire diameters, to permit a low reverse mu and good transfer characteristics.

The angle at which the electrons at the beam edges enter the decelerating field beyond the screen-grid plane can be expressed

$$\tan \alpha 0 = \frac{b}{2 f_a} \sqrt{\frac{V_1}{V_2}} \tag{16}$$

where b is the grid pitch. The potential terms may again be removed to give

$$\tan \alpha_0 = \frac{\left(\frac{\mu a}{s} - \frac{1}{3} \right)}{4 \frac{a}{b}} \sqrt{\frac{1}{1 + \mu + (s/a)}} \tag{17}$$

The aberrations in the beam can be approximated by considering the variation in mu over the cathode surface. If the cathode is divided into small strips over which the amplification factor is essentially constant, the focal length for each pair of axisymmetric strips can be calculated by repeated application of the focal-length equation, using the amplification factor associated with each

pair of strips. An example of the degree to which the calculated beam shape approximates the actual beam shape may be seen by comparing Figs. 12 and 13. The average focal lengths of the calculated beams are in relatively good agreement with those found in the traced beam, but the aberrations are similar only in the second case where a/b = 0.50. However, the extreme cases where a/b = 0.33 and 1.00 are not normally encountered in tube designs, and the 0.50 value is near the compromise usually made. The entrance angles of the calculated beam envelopes are considerably larger than those of the actual envelopes. This fact is due, primarily, to the thin lens assumption. Calculation of the beam shape with equations as limited as those presented in this text should be considered a guide only. An accurate determination requires a considerably more exact mathematical treatment or, better yet, a field plot of the structure upon which trajectories may be traced.

DIRECTLY INTERCEPTED SCREEN-GRID CURRENT

When the plate potential of a beam power tube is higher than the screen-grid potential and the control-grid potential is between the conditions of cutoff and control-grid current, the screen-grid current is caused by electrons which are directly intercepted as the electron rays pass through the screen-grid aperture. Many high-perveance beam tubes derive the majority of this directly intercepted screen-grid current from electrons which do not leave the cathode normal to its surface. This current interception cannot be accurately predicted, in the general case, except by ray tracing through the electrostatic field. The mathematical treatment of the beam shape, however, does suggest that:

1. The directly intercepted current should be at its minimum when the minimum beam cross section occurs in (or near) the plane of the screen grid.
2. The amount of current intercepted will depend upon the current distribution at the cathode surface.

If a control-grid potential is chosen such that the minimum beam cross section occurs beyond the screen grid and the control grid is then driven more negative, one would expect the percentage of the cathode current going to the screen grid to pass through a minimum and then increase as the grid becomes more negative. If the field is not constant over the cathode, the percentage of the cathode current going to the screen grid will go through a maximum with a further decrease in control-grid potential and then decrease again, because the current flow is cut off at the edge of the beam before it is cut off at the center. Fig. 14 is a plot of the direct screen-grid-current interception as a function of the ratio of control-grid potential to cutoff potential E_{c1}/E_{c0} in three different tube structures.* The amplitude differences between maxima and minima are predicted by the a/b ratio of each structure, except for the structure marked "III," where the minimum beam cross section is displaced only a small amount from the screen-grid plane before cutoff under the control-grid wires occurs. The location of the minimum beam cross section, when the control-grid potential is the same as the cathode potential, is as follows for each structure:

Structure I: a relatively large distance before the screen-grid plane.
Structure II: a small distance before the screen-grid plane.
Structure III: in the screen-grid plane.

It can be concluded that the direct screen-grid-current interception over the operating range can be held to a minimum for a particular structure (whose control grid is not driven positive) by designing it so that,

1. the minimum beam cross section occurs in the screen-grid plane when the cathode and control-grid potentials are the same.
2. a/b is small enough to restrict the formation of a sizeable maximum in the negative control-grid region.

The minimum beam cross section will occur in the screen-grid plane when the electrons leaving the cathode from the region under a control-grid wire converge to a line focus slightly before the screen-grid plane. The a/b ratio chosen, of course, should be a compromise to best suit the requirements of a reasonably sharp cutoff transfer characteristic and desirable beam shape, as well as acceptably low direct screen-grid-current interception.

HIGH VOLTAGE CONSIDERATIONS[7, 8, 9, 10]

Some Observations on the Breakdown Phenomena

When beam power tubes are used in vertical- and horizontal-deflection service, high-voltage performance must be considered in the tube design. The tube must withstand high-voltage pulses which appear on its plate, without "breaking down." A breakdown may be considered the result of a change in the electrical circuit between the two or more electrodes of concern. For example, when the potential difference between two "vacuum-spaced" electrodes is increased, a gradient is eventually reached at which field-emission effects are observed. For many smooth metal surfaces not specially processed to remove small surface irregularities or impurities, this gradient is of the order of 100 kilovolts per centimeter.[8] The emission-current density may be relatively high when the current flows between little points on the surfaces, to the extent that excessive localized heating of the bombarded electrode occurs. Gas is then driven from these hot points, and, if the pressure is high enough, arcs will form. When the phenomenon is first observed, these discharges are likely to be intermittent, and may even "clean up" as points are burned away. However, if the gas concentration is too high or further potential increases are too large or rapid, the vacuum will no longer be "self-healing" and a sustained arc will result.

The discharge along the surface of an insulator appears to be a more complicated phenomenon. At low enough potentials, the current is determined by the potential difference across the insulator and the resistance between the electrodes. As the surface gradient

*First described by O. Schade Sr., Ref. 1, p. 167.

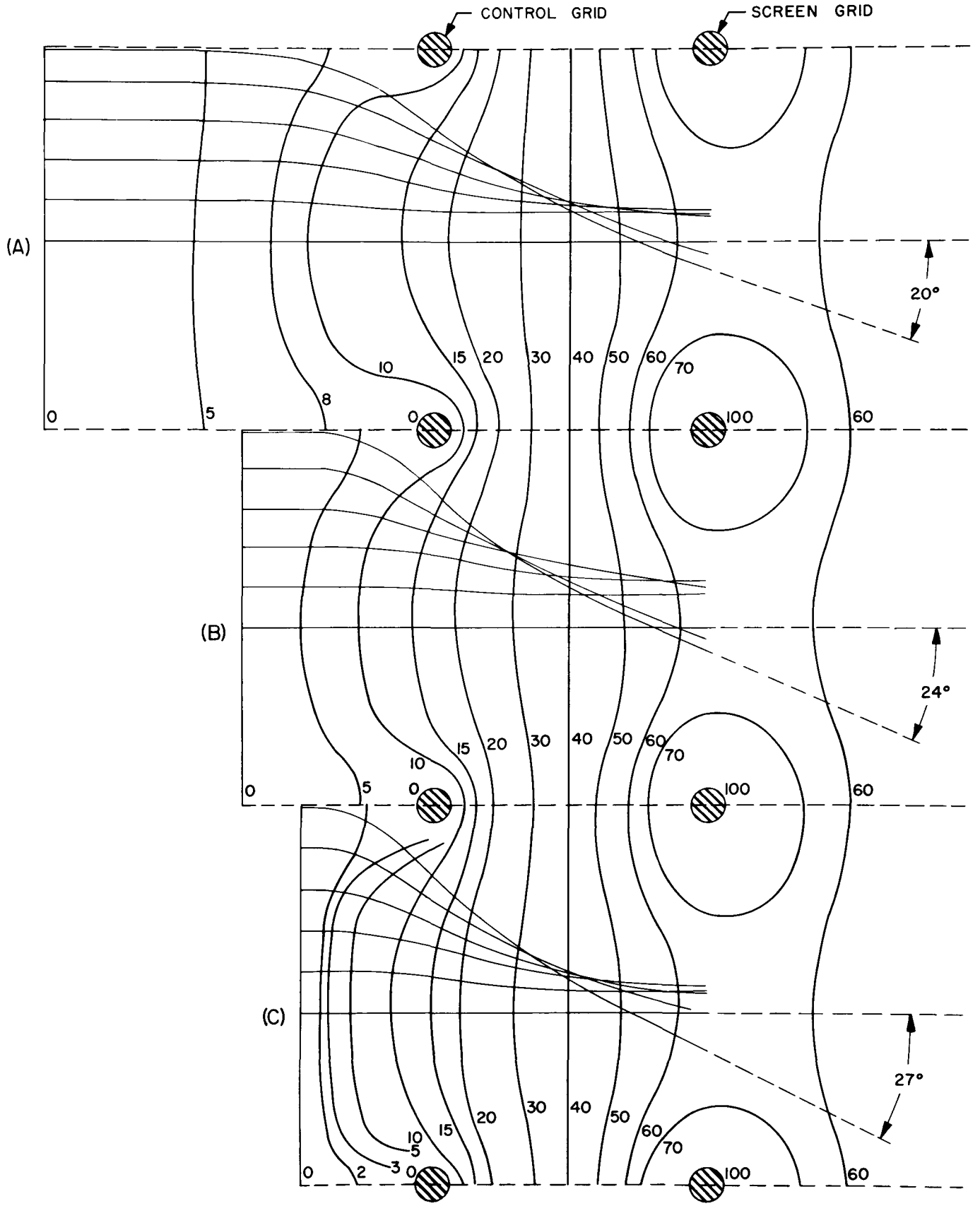


Figure 12. Pitch-to-Spacing-Ratio Effects on Beam Shape: (A) a/b = 1.00; (B) a/b = 0.50; (C) a/b = 0.33

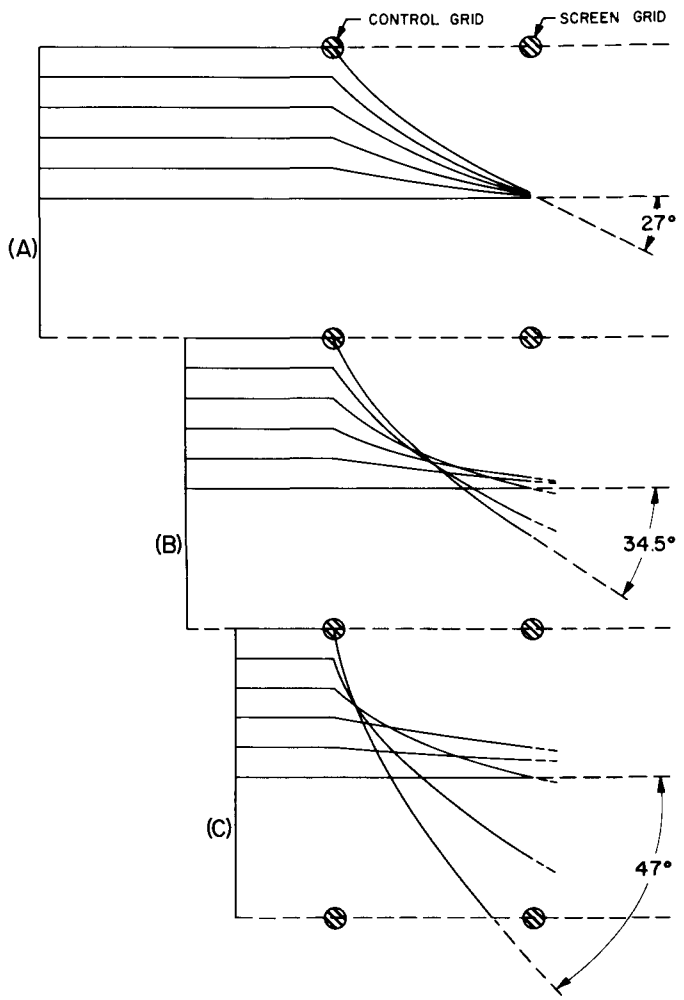


Figure 13. Calculated Beam Shapes: (A) a/b = 1.00; (B) a/b = 0.50; (C) a/b = 0.33

is increased, intermittent discharges develop. These might be local discharges on the surface, perhaps at areas or irregularities of particularly high gradient, or they might be discharges through a gas plasma near the surface which shunts the normal surface resistance. A further increase in gradient eventually results in a continuous discharge. This discharge may have a plasma associated with it; for discharges across mica, it is not uncommon to see an arc jump across a leakage slot. A fluorescence can sometimes be seen near the junction of insulator and electrode. The phenomenon is further complicated by the surface condition of the insulator. For example, the mica may be covered with alumina, getter flash, carbon, and metals from the tube electrodes. In addition, gases from the electrodes and insulator are present.

Some of the design features which permit good high-voltage characteristics will now be mentioned. There are limitations to the ability of conventional structures to withstand high potentials. Tubes with mica electrode spacers will not perform satisfactorily at potentials above 10 kilovolts unless precautions are taken to radically lengthen leakage paths along the mica surface. Tubes should be "double-ended" when potentials above 5 kilovolts are used, so that breakdowns or excessive leakages do not occur at the base-pins or the tube socket. Getter-flash position may also be an important factor.[7]

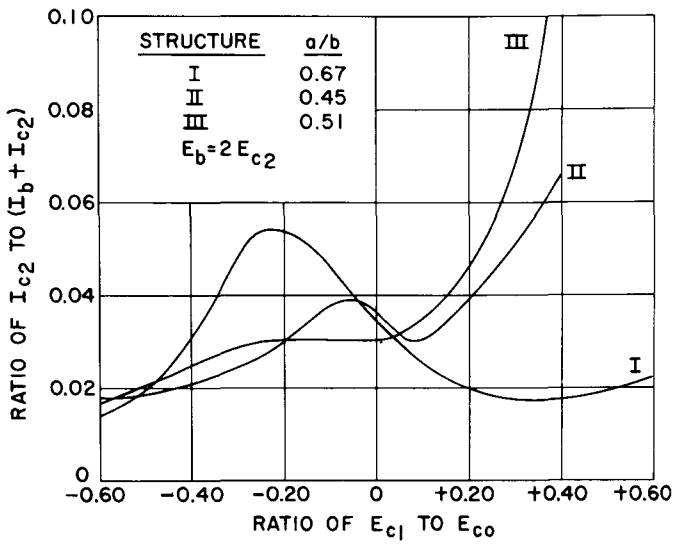


Figure 14. Directly Intercepted Screen-Grid Current as a Function of Control-Grid Potential

Considerations for Breakdown Between Metal Electrodes

To minimize the possibility of breakdown between metal electrodes:

1. The surfaces should be as smooth as possible. Surfaces such as those found on carbonized plates are inferior to clean, polished nickel. The aluminum-clad-steel surface generally gives much less trouble in horizontal-deflection tubes than carbonized plate surface.
2. The electrodes should be as cool and gas-free as practical.
3. The geometry of the electrodes should be chosen so that voltage gradients are kept low, definitely below 100 kilovolts per centimeter, unless unusual care is excercised to provide exceptionally smooth, clean surfaces, such as those produced by electropolishing. A gradient of 70 kilovolts per centimeter has been used successfully. It may be necessary to round off the edges of the beam-confining plates.
4. In marginal situations, the tube should be spot-knocked. Spot-knocking is the process of applying a high voltage from a source of moderate power capabilities to remove electrode and insulator surface irregularities. It is necessary to spot-knock at potentials above the tube rating (in some cases 30 to 40 per cent higher) to derive any real benefit from the process. A moderate series resistance in the high-voltage lead will reduce the severity of any discharges that occur. The value of the resistor should be determined by experiment; values of 50 kilohms to 500 kilohms are typical.

Considerations for Breakdown Along Insulator Surfaces

To minimize the possibility of breakdown along insulator surfaces:

1. Keep the voltage gradients low. Clean mica will withstand surface gradient of about 50 kilovolts per centimeter. Mica coatings such as alumina can increase this figure to about 80 kilovolts per centimeter. The approximate surface gradients can be determined by field plots in the electrolytic tank or by the resistance-paper field analogy. Safety-factors to both these limits should be considered.
2. Keep clearances between the mica and the electrodes it is supporting great enough to keep the gradient below 50 kilovolts per centimeter.
3. Make leakage slots wide enough to keep the gradient across them below 50 kilovolts per centimeter. Some leakage slots must hold off almost the entire electrode potential difference. This condition can be quickly seen from a field plot.
4. Keep the mica surface free from metallic deposits.
5. Consider spot-knocking.

THE MECHANICAL-THERMAL-ELECTRICAL DESIGN OF BEAM TUBES[1, 11, 12, 13]

Design Procedure for the "Optimum" Tube

A design procedure will be presented in outline form. Many other procedures are suitable. The choice of procedure depends upon the extent to which design work is required and upon the criteria imposed.

1. Determine the cathode area required. Factors of heater power and permissible current densities must be considered. It is advantageous to use the highest permissible current density if an efficient design is required. Direct current drains up to 100 milliamperes per square centimeter are presently feasible.
2. Determine the style of the structure to be used. The conventional, frame-grid, and multi-siderod structures are discussed briefly at the end of this article. Practical limitations of wire size and span, and mechanical quality of the grids must be considered when the cathode geometry is chosen.
3. Pick the smallest grid-to-cathode spacing that is practical for the application. This permits a choice of the smallest a/b ratio acceptable for the transfer characteristic.
4. Choose a control-grid wire diameter—a small one is desirable.
5. The materials used in the grid and supporting structure should have the best conductivity and lowest thermal emissivity possible under the limitations of the fabrication techniques and the conditions under which the materials are to be processed and operated.
6. Calculate the average control-grid temperature. A limit of 300 C is sufficient for many tube applications, but higher-than-normal grid-circuit impedances may require a lower limit.
7. If the control-grid temperature is below 300 C, the grid-to-cathode spacing may be decreased and smaller wires may be used, provided that this is mechanically feasible.
8. If the control-grid temperature is above 300 C, consider the possibility of reducing the thermal resistance from the grid wires to the heat sink (normally the chassis). If this value cannot be reduced there is no recourse but to increase the grid-to-cathode spacing and wire size until an acceptable temperature is reached.
9. When the cathode size, grid-cathode spacing, pitch, and wire size have been determined, the plane of the screen-grid wires may be chosen. The resulting triode amplification factor and beam shape should then be determined. If the results are not acceptable, changes should be made in the design procedure
10. Next, choose the screen-grid wire size. It is usually advantageous to use screen-grid wires which are larger than those of the control grid. Large wires are better heat conductors, impress larger potentials in the screen-grid plane, are easier to handle, and provide higher plate resistance and better high-voltage cutoff characteristics than small wires, although they will intercept a little more cathode current. Wires three times the diameter of the control-grid wires have been used successfully, when the beam shape is good.
11. The grid design is completed by determining the average screen-grid temperature. Audio applications may permit temperatures of 500 to 550 C, but horizontal-output service (with high-impedance screen-grid circuitry) requires a limit in the order of 400 to 450 C to assure that screen-grid emission does not become excessive.
12. The plate position is determined by the desired suppression characteristics. The cooling should be sufficient to limit temperatures to 450 to 500 C. Beam-confining-plate geometries are also chosen now, and high-voltage requirements considered, if necessary.

Some General Characteristics of Tube Structures

The Conventional Structure. The mechanics of this conventional structure are such that grid-to-cathode spacing and grid-wire size are near their limits. Where high-conductivity materials such as molybdenum, tungsten, and copper are used in the grid circuits, temperature rises in the stem leads, lateral wires, and siderods are of the same order. Grid radiators are reasonably effective if made large enough and of low enough thermal resistance. Cathode-current densities in the high-perveance tubes are relatively low, indicating a more-or-less brute-force approach to better performance. Plate temperatures, grid temperatures, and high voltage ratings have little or no safety factor.

The Frame-Grid Structure. The frame-grid structure being considered for beam power tubes consists essentially of a rigid, rectangular frame similar in outline to a picture frame. The lateral wires may be attached to the frame by welding or brazing, or with the conventional "notch-and-peen" technique, as used in Sylvania Electric Products' "framelock" design. Usually, an attempt is made to produce tension in the lateral wires.

Present techniques, primarily for economic reasons, do not permit the use of high-conductivity materials throughout the grid circuits. Grid lateral-wire temperature rises can become considerably larger than those of the supporting frames and stem leads unless the wire span is kept small. Radiators are much less effective than in the conventional structure. Performance can be

improved and current densities increased. No marked improvement in dissipation or high-voltage ratings is realized, but the feature of "automatic alignment" of grid wires may represent a cost advantage over the conventional structure.

The Multisiderod, Cylindrical Structure. A "Nuvistor" beam power tube with multisiderod, cylindrical grids is now being developed. A helical wrapping of "lateral" wire is brazed to the grid siderods. The relatively small span of lateral wire permits the use of small wire diameter throughout the grid without a loss of mechanical strength.

Although the techniques for making beam power tubes in the multisiderod cylindrical structure have not been fully developed, the mechanics appear to offer some definite advantages. If ceramics are used for the stem and enclosure, the plate circuit can be built with exceptionally low thermal resistance, allowing unusually high plate dissipations in small tubes. There is practically no temperature rise in the lateral wires, but large rises are found in the siderods unless they are kept quite short. For this reason, cathodes become short and large in diameter, and heater end-losses must be given consideration. High current densities and good performance can be attained. In addition, the structure is well suited to automatic assembly.

LIST OF SYMBOLS

E	Field potential gradient
E_1	Field potential gradient before the control-grid aperture
E_2	Field potential gradient beyond the control-grid aperture
E_b	Plate potential
E_{b0}	Knee plate potential
E_{c0}	Control-grid cutoff potential
E_{c1}	Applied control-grid potential
E_{c2}	Applied screen-grid potential
V_1	Potential in the control-grid plane
V_2	Potential in the screen-grid plane
V_{min}	Space-charge minimum potential
J	Current Density
a	Control-grid-to-cathode spacing
b	Grid pitch
s	Control-grid-to-screen-grid spacing
X_0	Screen-grid-to-virtual-cathode spacing at saturation
x, x_m	Distances
y, y_m	Distances
f_a	Apparent focal length
f_0	Corrected focal length
α	Entrance angle to the normal at a field boundary
α_0	Entrance angle of electron at screen-grid plane
α_L	Limiting entrance angle at screen-grid plane
ρ	Ratio of plate-to-screen-grid spacing to screen-grid-to-cathode spacing
δ	Distance normalized to X_0
μ	Amplification factor
τ	Normalized current transmission through the virtual cathode
ϕ	Ratio of plate potential to screen-grid-plane potential

REFERENCES

1. Schade, O. H., Sr., "Beam Power Tubes," Proc. of the I. R. E., Feb. 1938, Vol. 26, pp 137-181
2. Spangenberg, K. R., Vacuum Tubes, McGraw-Hill, 1948, p 248.
3. Beck, A. H. W., Thermionic Valves, Cambridge Univ. Press, 1953, p 180, 295.
4. Code F1
5. Davisson, C. J. and C. J. Calbick, "Electron Lenses," Phys. Rev., 1931, Vol. 38, p 585, and 1932, Vol. 42, p 580.
6. Bull, C. S., "The Alignment of Grids in Thermionic Valves," J. I. E. E., 1945, Vol. 92, part III, p 86.
7. Code F2
8. Millikan, R. A. and B. E. Shackelford, "On the Possibility of Pulling Electrons from Metals by Powerful Electric Fields," (abstract), Phys. Rev., Vol. 15, 1920, p 239.
9. Code F3
10. Code F4
11. Code F5
12. Code F6
13. Schade, O. H., Jr., "Heat-Transfer in Receiving Tubes," Electron Tube Design, RCA, (this text).

Audio Output, Vertical-Deflection, and Horizontal-Deflection Tubes for Receiving Applications

M. Bondy

Harrison

The purpose of this chapter is to present a practical approach to the design of receiving-type power tubes for such important uses as the output stages of audio, video, vertical-deflection, and horizontal-deflection circuits.

AUDIO OUTPUT TUBES

DESIGN OBJECTIVES

The first step in the design of an audio output tube is to establish the design objectives. These objectives are:

Power Output

Depending on the application, an audio output stage of one or more power tubes may be expected to deliver at the secondary of the loudspeaker transformer anywhere from a few milliwatts for a personal-type receiver to as much as 50 watts for the highest-quality home systems. One and one-quarter watts is normally considered adequate for table-type radios and television receivers. Three to five watts is normally needed for auto radios and better quality home receivers. For high-quality auto radio, 10 to 15 watts is required. For highest-quality home-entertainment equipment, 10 to 50 watts must be available, usually from a pair of tubes operated in a push-pull stage.

Power Sensitivity

This term is defined as the quotient of power output P_o in watts divided by the square of the grid signal E in root-mean-square (rms) volts required to produce that power output. It may be written as P_o/E^2. The grid signal required at the output tube to produce the chosen power output determines the voltage gain required between the audio input and the grid of the power output tube. For low-cost receivers, where keeping down the number of stage or improving their gain performance is more important than achieving low distortion or high peak-power-handling capability, the power sensitivity is especially important. For higher-quality receivers, the designer must consider not only power sensitivity, but must also give special attention to achieving low distortion and ample peak-power-handling capability.

Distortion

For receivers employing a single power output tube, total harmonic distortion up to 10 per cent is normally acceptable at maximum power output. Most of this total should be due to second-harmonic distortion because a third-harmonic component of only a few per cent sounds dissonant.

In the case of high-quality receivers, a total distortion of one to two per cent (or even less than one per cent) is the rule rather than the exception. For such receivers, almost all of the distortion results from odd-harmonic components because the prevalent use of push-pull output stages causes the even-harmonic components to cancel out.

Plate-Circuit Efficiency

The amount of dc power required to supply the plate or plates of the output tube(s), and particularly the amount of dc current, is of prime interest to the circuit designer because of cost considerations. Given a certain power requirement, the equipment designer usually chooses the circuit and tube which calls for the lowest plate-current drain. Plate-circuit efficiency N is defined as the ratio of audio power output to plate input and is usually expressed as a percentage, thus:

$$N = \frac{\text{Power Output}}{E_b I_b} \times 100$$

Knowing the power output desired and the efficiency of various circuits and tube types, the tube designer is ready to start his design. His first step is to decide which class of tube—triode, pentode, or beam power tube—offers the best choice. A discussion of the possibilities follows.

DESIGN CONSIDERATIONS

Whether the power tube design calls for a triode or a pentode depends on the application. Each class of tube has its advantages and its disadvantages so that a careful analysis of cost and performance capabilities is required. The following paragraphs discuss these matters by class of tube.

Triodes

A power triode produces its highest positive-swing value of maximum-signal plate current I_{max} at a relatively high plate voltage and its lowest negative-swing value of maximum-signal current I_{min} at a relatively

low plate voltage. The plate-circuit efficiency of a power triode is therefore low and the problem is to make I_{max} as large as possible and I_{min} as small as possible in order to obtain high output. The following formulas* for a power triode operated in a class A amplifier circuit can be derived from Fig. 1.

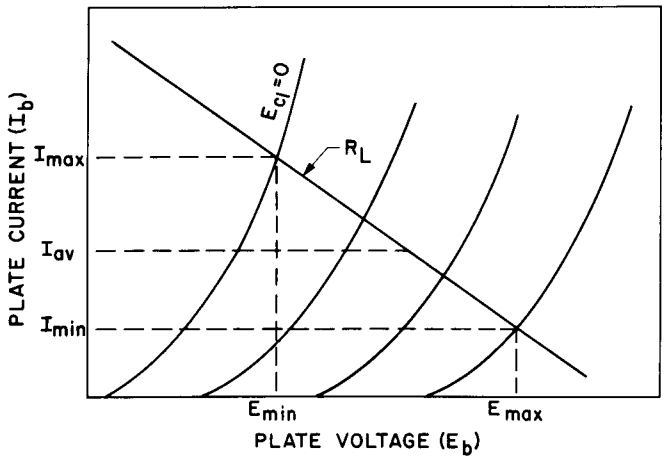


Figure 1. Typical Load Line for Triodes Used as Class A Power Amplifiers, Showing Critical Parameters

$$\text{Power Output } (P_o) = \frac{(I_{max}-I_{min})\,(E_{max}-E_{min})}{8} \quad (1)$$

Percent 2nd-Harmonic Distortion =

$$\frac{I_{max} + I_{min} - 2\,I_{bo}}{2\,(I_{max}-I_{min})} \times 100 \quad (2)$$

Plate-Circuit Efficiency (N as per cent) =

$$\frac{\text{Power Output}}{E_b\,I_b} \times 100 \quad (3)$$

$$\text{Load Resistance } (R_L) = \frac{E_{max}-E_{min}}{I_{max}-I_{min}} \quad (4)$$

Units should be given in volts, amperes, ohms, and watts.

With a triode design, the following comments are of interest:

In order to achieve a large value of I_{max} from a triode at minimum plate voltage, it is preferable to design for a low amplification factor.

In order to keep the amplification factor μ at a reasonable value (too low may mean an excessively high grid-signal requirement), it is necessary to increase the size of the cathode to obtain the same peak current at a low plate voltage.

Power sensitivity of a receiving-type power triode is usually very poor.

The distortion of a power triode is primarily due to the second-harmonic component. Therefore, where triodes are used in a push-pull circuit, the total distortion is very low because even-harmonic components are canceled out.

Plate-circuit efficiency of triodes is not as good as that of pentodes. For a triode, the minimum negative-swing plate current at maximum output is usually about 10 per cent of the zero-signal plate current, and the minimum negative-swing plate voltage is usually only about 50 per cent of the zero-signal plate voltage. Plate-circuit efficiency of a triode in class A operation, therefore, is only about 20 to 25 per cent.

Pentodes

The following formulas* apply to power pentodes (and also to beam power tubes) operated in class A amplifier circuits. Refer to Fig. 2.

Power Output (P_o) =

$$\frac{[(I_{max}-I_{min} - 1.41\,(I_x - I_y)]^2\,R_L}{32} \quad (5)$$

$$\text{Load Resistance } (R_L) = \frac{E_{max}-E_{min}}{I_{max}-I_{min}} \quad (6)$$

Per Cent Second-Harmonic Distortion =

$$\frac{I_{max} - I_{min} - 2I_b}{I_{max} - I_{min} - 1.41\,(I_x-I_y)} \times 100 \quad (7)$$

Per Cent Third-Harmonic Distortion =

$$\frac{I_{max} - I_{min} - 1.41\,(I_x - I_y)}{I_{max} - I_{min} - 1.41\,(I_x - I_y)} \times 100 \quad (8)$$

where I_x = plate current in amperes at $0.293\,E_{g1}$, and
I_y = plate current in amperes at $1.707\,E_{g1}$, and
with E_{g1} = max. peak grind signal in volts.

With a pentode design, the following comments are of interest:

Maximum undistorted power output is obtained from a pentode when it is operated so that the load line drawn on the plate-family graph intersects the zero-bias voltage curve at the knee. A pentode is a form of square-law amplifier.

When a pentode is operated with signal so that the intercept on the load line between average plate-current value and the maximum current value is equal to the intercept between the average value and the minimum

*Terman, Radio Engineering, McGraw-Hill

value, the second-harmonic distortion will be low. The third-harmonic distortion will usually be fairly high, i.e., about five to ten per cent. The use of a lower value of load resistance will reduce the third-harmonic component, but at the expense of reduced power output and increased second-harmonic distortion.

Because the screen-grid voltage is essentially constant regardless of the amount of plate current that flows, and because the plate will still collect current when its voltage is below that of the screen grid, a pentode designed with a much smaller cathode than would be required for a triode of the same power output can give higher plate current than a triode at low plate voltage.

The power sensitivity of a pentode can be made very high. A pentode can have a high mu-factor (determined by the triode amplification factor of the cathode, control grid, and screen grid) and still supply high values of peak plate current.

A pentode has the limitation that a substantial part of the cathode current under signal conditions is collected by the screen grid rather than by the plate, and that the ratio of plate current to screen-grid current decreases with signal increase. In effect, this limitation determines the amount by which the plate voltage can be swung below the screen-grid voltage under signal conditions. The ratio of plate current to screen-grid current for a pentode with signal can range between three to one and eight to one, and depends on tube design, particularly, the tube amplification factor. For example, a typical pentode, which has a ratio of average plate current to the average screen-grid current for zero-signal conditions of six to one, will show a ratio at maximum signal of only three and one-half to one because screen-grid current rises during the downward sweep of the plate voltage. For this example, if the desired output is four watts and the plate-circuit efficiency is 40 per cent, a plate input rating of 10 watts (four divided by four-tenths) is called for. Correspondingly, if plate and screen-grid voltages are the same, the screen grid must be designed to have a dissipation rating of 2.85 watts (10/3.5, where 10 is the plate dissipation rating under zero-signal conditions and 3.5 is the ratio of plate current to screen-grid current at maximum signal). Because this screen-grid dissipation rating is fairly high, the tube design would probably require a screen-grid structure utilizing high-conductivity lateral wires and siderods.

Beam Power Tubes

The formulas shown for pentodes used in class A amplifiers also apply to beam power tubes. The following comments will be of interest:

For beam power tubes, the plate family of characteristics curves shows the knees of the curves as quite sharp as compared with those of a pentode. The load line, therefore, should not cut below the knee of the zero-bias-voltage curve. This method of operation tends to produce second-harmonic distortion. When beam power tubes are used in push-pull circuits, the even-harmonic components cancel out, and high output is produced at very low distortion.

Beam power tubes are usually designed to have a lower mu-factor than pentodes, because a relatively low mu-factor is necessary to achieve proper beam formation. As a result, the power sensitivity of a beam power tube is not as high as that of a pentode.

Screen-grid dissipation is much less of a problem with beam power tubes than with pentodes. Ratios of average plate current with zero signal to average screen-grid current with zero signal of ten to one or even twenty to one are usual for beam power tubes. Ratios under conditions of maximum signal ranging between seven to one and fifteen to one can be expected. Thus, a beam power tube which is required to produce four watts output at 40 per cent plate-circuit efficiency (plate input rating of 10 watts) and which has a seven to one ratio of plate current to screen-grid current at maximum signal would need to have a screen-grid dissipation of only 1.42 watts, as compared to 2.85 watts for a pentode having the same power-handling capability. It is possible, therefore, to use the pentode cathode in a beam power tube and with it to obtain increased output (limited only by the plate dissipation capability of the beam power tube).

DESIGN EXAMPLE

An equipment designer who was using the 6AQ5 beam power tube in a TV receiver was concerned about the power output that he obtained when his set was used in weak-signal areas. Normal operating conditions were:

1. 215-volt B supply.
2. Grid-bias voltage obtained from a cathode resistor operating at a total cathode current of 0.035 ampere.
3. In good-signal areas, the grid signal to the tube was over nine volts rms and the output was over two watts.
4. In weak-signal areas, the grid signal to the tube dropped to five volts rms and the output fell below one watt.

Objectives of Design Problem

The equipment designer desired to use the same transformer (resistance load equal to 5000 ohms), with no change in plate-current drain, to obtain an output of not less than 1.25 watts with a five-volt rms signal, and of over two watts with a nine-volt rms signal.

Analysis of Design Problem

The existing circuit design was analyzed to determine its capabilities. The actual plate voltage applied to plate and screen grid was equal to the plate-supply voltage minus the voltage drop in the cathode resistor. With a total current drain of 0.035 ampere and the knowledge that the ratio of plate current to screen-grid current of the 6AQ5 at maximum output is seven to one, it can be estimated that the plate current is 0.030 ampere. Fig. 2, which is the plate family of curves for the 6AQ5 with a screen-grid voltage of 200 volts, shows that the grid-bias voltage for 0.030 ampere is -13 volts. Thus, the actual plate voltage is 215-13 or 202 volts.

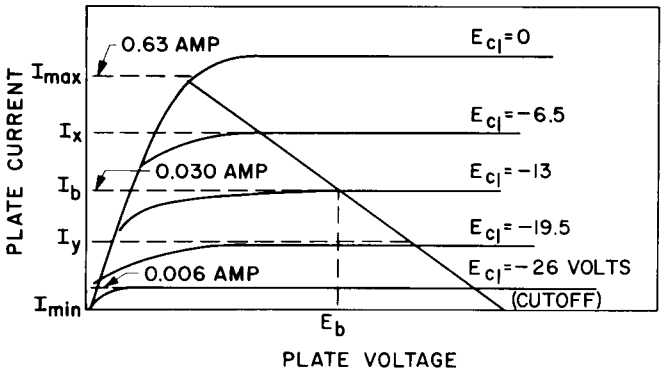


Figure 2. Plate Characteristic of 6AQ5 with Screen Voltage of 200 Volts. Typical Load Line Is Shown When Used as a Class A Power Amplifier. Critical Parameters Are Shown

By means of the load line and Eq. (1), an estimate of the power output gives:

$$P_O = \frac{(0.063 - 0.006)\ (325 - 35)}{8} = 2.07 \text{ watts}$$

and for a weak signal of five volts rms (seven volts peak)

$$P_O = \frac{(0.051 - 0.013)\ (285 - 95)}{8} = 0.9 \text{ watts}$$

In order to obtain the desired improvement in this receiver, it is necessary that:

1. The new tube design have a plate family with knee current at least as good as that of the 6AQ5, or with a lower knee voltage.
2. The new design have higher power sensitivity. This feature can be obtained by designing the new tube to have smaller grid-bias voltage, higher mu-factor, and higher transconductance. For example, if the new design had exactly the same plate family, except that the control-grid voltages were reduced to 80 per cent of the present values (i.e., in the ratio of one watt to 1.25 watts), then a four-volt rms signal (5 x 0.8 = 4) would give 0.9 watts output. For a five-volt rms signal, the new power output can be approximated as follows:

$$P_O = \frac{5^2}{4^2} \times 0.9 = 1.4 \text{ watts}$$

3. It can be assumed, because distortion was not mentioned, that distortion no higher than that of the 6AQ5 is acceptable. Lacking such information it could be assumed that the accepted figure of up to 10 per cent would be satisfactory as a starting point.

Possible Solutions

A relatively simple solution, if it will give the desired results, is to use the 6AQ5 structure but to redesign the grids. One way to increase the mu-factor is to increase the turns per inch of the control grid. In fact, beam-forming considerations usually require that the increase be done this way rather than by increasing the spacing between the control grid and the screen grid. The turns per inch of the screen grid would also have to be increased to that of the control grid because beam power tubes use aligned control grids and screen grids. The process of aligning grids requires that both grids have the same turns per inch and be adjusted so that each lateral wire of the screen grid is shaded from the cathode by a lateral wire of the control grid. The use of aligned grids is one factor which provides beam power tubes with a high ratio of plate current to screen-grid current. Because an increase in mu-factor will lower the plate current at the knee of the zero-bias curve, the control grid will have to be moved nearer to the cathode to raise the perveance of the tube. If this change makes the spacing between control grid and cathode too small for practical manufacture, the wire size of the control grid should be reduced to keep the spacing reasonable, and the turns per inch should be further increased to compensate for the smaller wire size. The turns per inch, wire size, and spacings of the modified design may have to be changed to obtain the best compromise between the desired electrical characteristics and practical mechanical design.

Assume that the new design data indicate that any design which would meet requirements would be extremely difficult to make. For example, the turns per inch for the grids might be too high, the wire size too small, or the electrode spacings too close for practical mass-production manufacture. With present-day manufacturing methods, it is very difficult to make aligned grids with turns per inch above 50, or with wire size below two mils. If the design data indicate that the beam power tube approach is not feasible, then the evaluation of a triode design would be in order.

As a first assumption, consider a triode to operate at a plate voltage of 200 volts and a plate current of 0.035 ampere, with a plate-circuit efficiency of 25 per cent at full ratings.

Then, plate input is equal to:

$$200 \times 0.035 = 7 \text{ watts}$$

and power output is equal to:

$$7 \times 0.25 = 1.75 \text{ watts}$$

Because the design requirement is two watts, this solution would not be successful. To use a triode for this application, it would be necessary to raise the input power to at least eight watts. With a plate voltage of 200 volts, the plate-current requirement would be at least 0.040 ampere.

The problem is to obtain from a triode design 1.25 watts at a five-volt rms signal and over two watts at a 9.2-volt rms signal. One method of attacking the problem is to construct a triode plate family (see Fig. 3) to fit the requirements. Normally, the load line is so chosen that the maximum plate current is equal to twice the average plate current. This rule usually results in a good compromise between power output and distortion. If distortion is too high, the load resistance must be increased. If distortion is low, the load resistance can be reduced with the benefit of increase in the output power. In this case it is not possible to draw a zero-

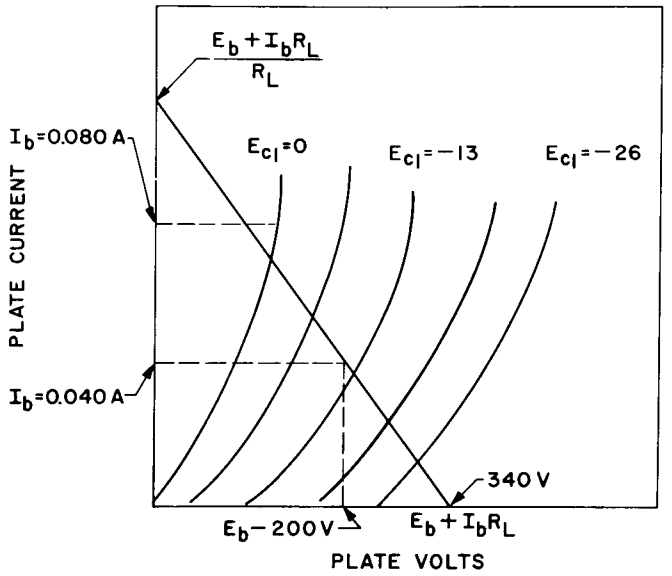


Figure 3. Reconstruction of a Triode Plate Family from Available Information. Any Number of Load Lines May Be Calculated. Theoretically, Any Family of Curves Can Be Drawn. Experience and Checks of Existing Triode Designs Must Be Used as Limiting Factors

bias curve which will go through the 0.080-ampere point. Either the plate voltage would have to be raised or the load resistance lowered to make it possible to locate the maximum current point as desired. A lower load resistance should be used. A load resistance of 3500 ohms has been selected, because this value results in a zero-bias voltage curve that looks fairly reasonable. The choice of load is based on a knowledge of existing power triodes. The point on the −26-volt curve can also be located. It should be as near cutoff as possible. This choice would make the amplification factor of the tube approximately 320/26 or 12, because amplification factor multiplied by twice the grid-bias voltage equals the peak plate voltage. Now, a rough calculation of power output shows that:

For a 13-volt peak signal, power output is approximately

$$\frac{(0.080 - 0.005)\ (320 - 60)}{8} = 2.4 \text{ watts}$$

For a seven-volt peak signal, power output is approximately

$$\frac{(0.060 - 0.025)\ (275 - 135)}{8} = 0.6 \text{ watts}$$

Without proceeding further, it is obvious that the power sensitivity would be poorer than that of the original design.

A new approach is to consider a triode design with a larger cathode. The formula for average plate current under signal conditions is:

$$I_b = \frac{KA}{d_{gk}^2}\left(\frac{\frac{E_b}{\mu} + E_{c1}}{1 + \frac{1}{\mu} + \frac{4}{3}\frac{1}{\mu}\frac{d_{gp}}{d_{gk}}}\right)^{3/2} \qquad (9)$$

where K = 2.33×10^{-6}
A = Area of Cathode (square inches)
d_{gk} = Grid to Cathode Spacing (mils)
E_b = Plate Voltage (volts)
E_{c1} = Grid Voltage (volts)
μ = Amplification Factor
d_{gp} = Grid-Plate Spacing (mils)

From this equation, it follows that the lower the triode amplification factor, the smaller the cathode can be for a given current level. Conversely, the higher the triode amplification factor, the larger the cathode should be for the same current capability.

A fairly straightforward method for determining the size of cathode needed for an improved triode design follows.

Assume a grid-cathode spacing. This assumption requires some understanding of the limitations of tube structures. No definite rules can be properly set up because better structures permit closer spacings of electrodes but currently, a grid with high turns per inch and small spacing between its siderods can have a grid-cathode spacing as small as 0.001 inch. For larger grids for use with flat cathodes, a grid-cathode spacing of 0.006 to 0.010 inch is reasonable.

Select a suitable ratio of grid pitch "a" (measured in thousandths of an inch between centers of adjacent grid laterals) to grid-cathode spacing b (measured in thousandths of an inch after tube processing). For a design which will have high plate current, good cutoff, and a high quotient of transconductance over plate current, a pitch-to-spacing ratio a/b of 1.5 is normal. For a design to give sharp cutoff, a ratio of 1.2 can be used.

The turns per inch and grid-cathode spacing for the desired tube design can now be located on the amplification factor nomogram (Vodges and Elder formula) in the article on the Electric Field and Space Current, by Scholz.

Choose a lateral grid-wire size and locate the ratio r/a (where r is radius of grid wire) on the graph to determine the grid-plate spacing.

If the values of grid-plate spacing or wire size prove to be unsatisfactory, modify the assumed values to obtain an acceptable compromise.

By means of the current value desired for the average plate current under signal conditions and Eq. (9), calculate cathode area A. Increasing the calculated cathode area by 25 per cent is suggested because the larger value gives better correlation between actual results and desired results.

From the area, compute the current density for the cathode with average quiescent plate cirrent. Normal-

ly, for power tubes, direct-current densities up to 0.080 ampere per square centimeter are used. However, some high-frequency tuner tubes use current densities as high as 0.250 ampere per square centimeter. The chief reason that power tubes are designed for relatively low cathode current densities is that design limitations imposed by mechanical considerations of grid-wire size, grid pitch, and grid-cathode spacing prevent the use of more efficient structures for this class of tube.

Carrying through on the triode approach would result in a manufacturable design. The cathode would, however, be very large, and the power required to heat it would be excessive — well beyond the point of practicality for the customer.

Another approach to the solution of the problem is to consider a pentode design. The analysis so far has shown that it is not practical to redesign the mechanical structure of the original 6AQ5 design and produce a design, using the original cathode, which would be suitable for mass-production methods. Therefore, a new design, whether it be a pentode or a beam power tube, must employ a larger cathode. A very practical design would be to design a pentode around a larger cathode and to employ nonaligned grids; in this case, only the control grid need use fairly high turns per inch.

Such a design would give costs close to, if not lower than, those of equivalent beam power tube design utilizing aligned control and screen grids with medium turns per inch and beam-confining plates, and requiring additional labor to align the grids. The choice between a pentode and beam power tube should be based on a careful cost analysis of each design. Consider the ratings desired of the new design. The equipment designer interested in an improved 6AQ5 can use ratings well within those of the 6AQ5. He needs a plate input of only 6 watts (0.030 x 200) and a power output of 2.5 watts. A pentode meeting these conditions and having a ratio of 3.5 for plate current to screen current under conditions of signal would need a screen-grid dissipation rating of 6/3.5, or 1.75 watts. However, a more versatile design, which would be of interest to 6AQ5 users, could be based on a 12-watt plate input power. Such a tube would require a screen-grid dissipation of 12/3.5, or 3.43 watts. This requirement would be considered high for normal receiving-tube pentode design, and would call for a special and expensive screen-grid structure. Therefore, an appropriate and more desirable design would be that of a beam power tube.

The objective for the new design is to obtain higher power sensitivity than provided by the 6AQ5, that is, the same power output with less grid drive. Fig. 4 shows the 6AQ5 and the desired characteristics. The mu factor is increased from 8 to approximately 12. A five-volt rms signal gives almost full power output. Comparative typical operating data are given in Table I for the 6AQ5 and the new beam power tube design.

The size of the cathode was calculated from Eq. (9) but was modified as follows:

Because the calculated size would have meant a new

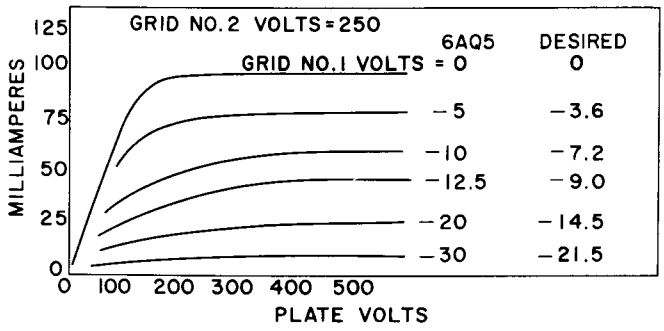


Figure 4. Plate Family of the 6AQ5 Showing the Desired Changes in Grid Voltages to Obtain the Desired Results

cathode design and, thus, new plate and screen-grid designs, the next larger size standard cathode was selected. New grid and spacing parameters were calculated to go with the new cathode. By this procedure, it was possible to use many standard parts and thus avoid the lost time and expense involved in new tooling.

After the cathode was chosen, the size of the plate and its configuration are determined on the basis of the desired plate-cathode spacing and the plate-dissipation rating. For receiving-type power tubes and plate materials, it is desirable to keep the plate dissipation below one watt per square centimeter. This limitation is not particularly critical, but too high a plate temperature may shorten tube life and make tube processing to achieve low gas content more difficult.

Critical Design Areas

Once the electrical characteristics have been estab-

Table I

Characteristic	6AQ5		Developmental Beam Power Tube	
Plate Volts	200	250	200	250
Screen-Grid Volts	200	250	200	250
Control-Grid Volts	-13	-12.5	7.5	9.0
Peak-Signal Volts	-13	-12.5	7.5	9.0
Zero-Signal Plate Ma.	30	45	30	45
Max.-Signal Plate Ma.	31	47	31	47
Zero-Signal Screen-Grid Ma.	3.0	4.5	3.0	4.5
Max.-Signal Screen-Grid Ma.	4.0	7.0	4.0	7.0
Power Output (watts)	2.1	4.5	2.2	4.5

lished, it is advisable to consider a few critical areas of the design. These are gas, grid emission, and audio distortion. They are discussed in the following paragraphs.

Processing of power tubes should always be based, first, on keeping the gas-current level as low as possible. The reason is that equipment designers will usually want to use the largest grid resistor permissible and, thus, will produce a high-impedance grid circuit which, with any reverse current due to grid emission or gas, tends to reduce the bias voltage appreciably and cause excessive plate current. Keeping the gas content low requires: first, that consideration be given to the choice of a plate configuration which can be uniformly heated to a high temperature during the exhaust cycle; and second, that materials for the heavy stamped parts be such that they contain only small amounts of occluded gas and can therefore be quickly degassed at the temperature levels normally produced by rf-heating equipment. Thick fluffy carbon coatings on steel base material contain much gas which requires a high exhaust temperature and a long exhaust cycle to degas the material properly. On the other hand, this material has excellent radiation properties, so that it runs cooler than other materials for a given power input. If the design involves a tube type which is critical for gas and which will be exhausted on production equipment providing a minimum of rf treatment, it might be better to choose a material such as aluminum-clad steel which, although it does not have the excellent radiation emissivity of the fluffy carbon-coated material, does have a much lower gas content. Such compromises in balance must be considered during the design of each tube type.

A second critical area is that involving grid emission which is also a reverse grid current. Either the control grid or the screen grid can show grid emission. In an audio power type of tube, screen-grid emission normally is unimportant, first, because the voltage gradient at the screen grid is low (the plate voltage rarely gets higher than twice the screen-grid voltage) and, second, because the impedance of the screen-grid circuit is normally so low that it would require a tremendous amount of screen-grid emission to cause the tube to operate poorly.

Control-grid emission, however, is an important phenomenon because, first, the voltage gradient at the control grid is high when the grid voltage is driving the tube into cutoff condition; second, the impedance of the control-grid circuit is high; and, third, the control-grid temperature is normally high enough, due to heat radiated from the cathode, to cause the control grid to emit. Actually, the screen grid runs as hot as the control grid (if not hotter) but, as mentioned, it is much less critical. The tube designer should take special care to make sure that his tube design shows no sign of grid emission. If there is evidence of any grid emission, then additional design precautions are necessary. These, arranged in an appropriate order, include:

First, reduce the grid temperature by improving the heat conduction from the grid to the stem leads. Use of higher-conductivity wire for grid laterals, siderods, and stem leads is a suitable method. Second, reduce the grid temperature by improving grid radiation. Because the grid is surrounded by electrodes, such as the cathode, screen grid, and plate, which usually operate hotter than the control grid, it is necessary to place any radiating structure in the clear above or below the mount structure. Such a radiator may help to reduce the grid temperature, but again, heat must be conducted from the grid to the radiator before it can become effective. Third, try surface plating the grid with metals such as gold, palladium, platinum, silver, and rhodium to reduce the activity of any barium which may deposit on the grid. Care must be taken to keep the metal plating from vaporizing during tube processing, or from combining with the grid material. Fourth, use a less-active cathode material to reduce barium evaporation from the cathode onto the control grid. Other methods include changes in processing procedures designed to boil off the evaporated barium from the grid; the use of a lower cathode temperature during processing; and the choice of an evaporant base matrix on the grid which will increase its work function.

A third critical area is that involving distortion. As soon as a design has been selected and tubes made, it is necessary to read many tubes for plate current at zero-bias-voltage and cutoff conditions, for distortion due to the second-harmonic component and the third-harmonic component, as well as for the total distortion. The results obtained will establish reference information for future use. Similar readings should be made on batches of tubes as soon as the type is put into regular production. This kind of information is useful for the following reasons: first, if these values are not controlled by the specifications, the values originally read supplement the normal control data and permit more accurate evaluation of the effects of future changes in tube parts or processing; second, such information can be used to evaluate a manufacturing problem such as "low power" to determine what went wrong and what should be brought back into line; and third, such information may be helpful in analyzing customer reports on tube performance. For example, a customer reported low power output on life for a well-established production type. His report was based on life test results of tubes operated in a push-pull circuit in which a pair of tubes was rejected if the pair failed to give a specified power output at 10 per cent distortion. The rejected tubes passed the production test for power output and distortion in a single-ended circuit, but further investigation revealed that recent tube production, due to a lower, more rounded knee on the zero-bias curves, was running higher in third-harmonic distortion and lower in second-harmonic distortion than the initial production. The differences, together with some minor emission slump on life, accounted for the low power output reported by the customer.

VERTICAL-DEFLECTION TUBES

DESIGN OBJECTIVES

The vertical-deflection circuit of a television set is basically a power output stage driven by a sawtooth voltage; it is required to produce a variable current which, when supplied to the vertical-deflecting coils of the

yoke, deflects the electron beam of the picture tube vertically (called the trace) in a linear manner, and then to return the beam quickly to its starting position (called the retrace). The entire process is called the scanning cycle. For simplicity, the deflecting coils can be considered as providing a resistive and inductive load in series with the output tube during the retrace portion of the scanning cycle, and as providing a pure resistive load during the trace portion. During the trace portion, the sawtooth signal drives the tube as a class A amplifier which supplies power to the yoke coils through a step-down transformer. During the trace portion, the vertical-deflection output tube requirements are very similar to those of an audio power amplifier tube. During the retrace portion, the similarity no longer exists, because the tube is subjected to high-voltage pulses of about 1000 to 2000 volts peak.

The power requirements of a vertical-deflection system depend upon the following factors: maximum deflection angle of picture tube, yoke design, neck diameter of picture tube, ultor voltage of picture tube, and losses of step-down transformer. Increasing any of these factors increases the amount of power which must be handled by the vertical-deflection output tube. For present-day picture tubes and circuitry, the power normally required will be one to six watts. Theoretically, the tube need handle only twice the required power (based on an ideal circuit efficiency of 50 per cent). In practice, the maximum efficiency attained for the sawtooth waveform is about 33 per cent for a triode, and about 40 per cent for a pentode.

DESIGN CONSIDERATIONS

Triodes

Triodes have been very popular as vertical-deflection output tubes. First, the characteristics of triodes are such that the step-down transformer can be designed to complement the transfer characteristics of a triode; the resultant current waveform is linear and may be obtained with good circuit efficiency. Second, triodes are usually lower in cost than pentodes or beam power tubes and can be incorporated in more-economical circuit designs. Third, in receivers which provide a boosted B-supply from the horizontal-deflection circuit, it is feasible to have a B-supply of as much as 550 volts. With this much voltage available, it is possible to design a triode having a fairly high amplification factor (about 16) and a small cathode. Such a tube would be easy to drive (have high power sensitivity) and, if rated at eight watts input to the plate, could deliver 2.6 watts at a circuit efficiency of 25 per cent. If less power were adequate, it might be desirable to place another medium-mu triode in the same envelope for use as the vertical-oscillator tube. Fourth, a triode may need a larger cathode than a pentode for a specified current at zero-bias voltage. Therefore, the triode can supply a specified current at a lower current density than a pentode, a condition which would normally mean that the triode would have longer life and more stable operation at low plate voltage than a comparable pentode.

The major disadvantages of triodes are associated with conditions of operation involving low B-supply voltages and the high values of current at zero-bias voltage and low plate voltage (more than 0.1 ampere at less than 100 volts) are needed, the cathode required for a triode must be very large, or the triode must have a low amplification factor. The situation is made worse by the fact that such high-current applications require the use of a resistor in the cathode circuit. The reason for such use is that large grid resistors are needed with high-impedance driving circuits and practical values of coupling capacitors. Because reverse grid current caused by gas, grid emission, or leakage in the tube can reduce grid bias voltage resulting from the use of a grid resistor and thereby permit excessive plate current, the cathode resistor regulates the plate current and protects the tube. Another method used in some applications for obtaining a degree of bias-voltage regulation employs feedback oscillator circuitry in which a discharge tube supplies the regulation voltage. When a low-mu triode is used, a cathode resistor of large value is required. The high-voltage drop of such a resistor decreases the plate-cathode voltage and thus reduces the power output capabilities of the tube. Furthermore, a low-mu triode calls for a large driving signal which may be difficult to provide. When the amplification factor of the tube is made high, a large cathode is required. Such a cathode may require so much heat that a double-triode design of tube is impractical. In addition, a large cathode increases the possibility of grid emission because the cathode may radiate excessive heat to the control grid.

Beam Power Tubes

Beam power tubes enjoy some popularity as vertical-deflection tubes because: first, beam power tubes require less grid drive than triodes; second, they can give higher power with a smaller cathode than an equivalent triode and, therefore, require less heater power; third, a pentode can be used with a minimum of circuit change to go from a design using a triode with boosted B-supply to a design not requiring a boosted B-supply, the latter design imposes less load on the horizontal-deflection tube; and fourth, beam power tubes permit operation with a fixed bias voltage and, thus, eliminate the need for a cathode resistor and its 60-cycle bypass capacitor.

Beam power tubes present certain disadvantages such as: need for damping components; need for special care in load-impedance match, even at times, to the extent of using yoke damping resistors with negative-temperature coefficients; good linearity is difficult to achieve and requires a certain amount of shaping of the grid signal; a screen-grid voltage-dropping resistor and bypass capacitor are needed when the screen grid is to operate at less than the plate-supply voltage.

DESIGN EXAMPLE

Choice of Tube

A new vertical-deflection output tube for operation from a 130-volt supply is needed. The first requirement is to determine the power which the tube has to supply. The amount of power needed depends on the ultor voltage of the picture tube, the deflection angle, and the overscan capability specified by the equipment

designer. Equipment designers may state their overscan requirement differently, as for example, 20 per cent overscan at centerline voltage or full scan at a line voltage of 100 volts, but allowance must be made for this factor in establishing the power requirements for the new vertical-deflection tube.

The power output of a vertical-deflection tube with sawtooth signal input is:*

$$P_o = \frac{(E_{max} - E_{min})(I_{max} - I_{min})}{12} \qquad (10)$$

If a power requirement of 3.7 watts and a plate-voltage swing of 200 volts for this power output are assumed for the new tube design, then based on a dc plate voltage of 130 volts, a knee voltage of 30 volts, and a peak plate voltage with signal of 230 volts, Eq. (10) can be used to determine the current swing for the required power output as follows:

$$I_{max} - I_{min} = \frac{P_o \times 12}{E_{max} - E_{min}} = 0.222 \text{ ampere}$$

To give good linearity with a step-down transformer having the proper matching impedance, the dc current for operating conditions should be 0.080 ampere.

A triode design which would be capable of giving over 0.2 ampere at low plate voltage and zero grid-bias voltage would require a very large cathode. Such a triode design would undoubtedly have to have a low amplification factor, would be hard to drive, would require high heater power, and would present grid-emission problems as basic disadvantages.

A pentode design looks more promising, but high screen-grid dissipation could present problems. A comparison of design objectives for a pentode and a beam power tube are shown in Table II.

It should be noted that the screen-grid-input objective and, therefore, the required screen-grid dissipation, is much higher than that of the beam power tube. Provided a screen-grid structure capable of handling the required power at a temperature below that which would cause grid emission can be manufactured on a reasonable cost basis, the pentode might have the advantage over the beam power tube in that it can be designed with a higher mu factor and, therefore, would require somewhat less driving signal. In addition, it might have a slight cost advantage. These advantages have to be weighed against the cost of making the pentode screen grid free from grid-emission problems and the probability that the beam power tube will have longer life expectancy due to cooler operation.

Table II

Characteristic	Pentode Design	Beam Power Tube Design
DC Plate Volts	130	130
Peak Positive-Pulse Plate Volts	1000	1000
DC Grid-No. 2 Volts	130	130
Peak Plate Ma.	260	260
Grid-No. 2 Ma. for Plate Volts = 30	90	30
Average Plate Ma.*	80	80
Average Grid-No. 2 Ma.*	20	7
Plate Input (watts)	11.0	11.0
Grid-No. 2 Input (watts)	4.0	2.0

*Measured at some grid-No. 1 voltage to be determined.

Critical Design Areas

So far as problems of grid emission, gas, and leakage between elements in vertical-deflection-output-tube designs are concerned, the discussion in this chapter on the same problems with audio power tubes applies. Problems peculiar to the vertical-output application include the following:

First, it is important to maintain consistency of the cutoff characteristics because of its effect on performance. Tubes having sharper cutoff characteristics than normal will cause the television picture to contract at the top of the picture-tube screen; tubes with cutoff characteristics more remote than normal will stretch the picture at the top of the screen. The linearity control on the TV receiver provides some correction but cannot compensate for wide differences in cutoff characteristics.

Second, the plate current at zero grid-bias voltage should be controlled to close tolerances, because this value determines the maximum picture height.

Third, grid temperatures should be low enough to insure conservative operating conditions. Control-grid temperatures should be less than 300 C; screen-grid temperatures should not exceed 500 to 550 C.

Fourth, the basing arrangement chosen should isolate the plate lead through the stem from leads operating at lower voltages in order to avoid stem leakage.

Fifth, the possibility of microphonism should be guarded against. Microphonism can produce many weird effects in a television picture.

*Rms value of triangular wave is $\frac{E_{max} - E_{min}}{2\sqrt{3}}$ for voltage and $\frac{I_{max} - I_{min}}{2\sqrt{3}}$ for current.

See Reference Data for Radio Engineers, 4th Edition, International Telephone and Telegraph, p. 1020

Sixth, the interference problems in a TV picture caused by high-voltage pulses must be considered. The article on "Circuit Troubles Caused by Unusual Tube Effects," by W. E. Babcock, discusses this problem.

HORIZONTAL-DEFLECTION TUBES

DESIGN OBJECTIVES

The horizontal-deflection circuit of a television set is a power output stage which is driven by a sawtooth grid voltage, and which is required to deliver current to the horizontal-deflecting coils in such a manner as to deflect the electron beam of the picture tube horizontally and linearly in one direction and then to return the beam quickly to its starting position. The tube acts essentially as the switch for a load which is primarily inductive.

For simplicity of analysis, consider the basic horizontal-deflection circuit as a tuned circuit and the deflection tube as a switch. (Refer to Fig. 5.) Fig. 6 shows the sawtooth grid-voltage output of the circuit. When the switch is closed at point 1 energy is built up in the magnetic field of the coil by the current, which increases exponentially to a peak value of I_1 at point 2. When the current reaches this point, the switch is opened. The energy stored in the LC circuit causes the circuit to oscillate at its natural frequency. After one-half of the natural frequency cycle has passed, the current reaches the negative peak value I_2 at point 3. At this point, the switch is closed again, the current decreases toward the zero line at point 4, and the circuit is ready to repeat the cycle.

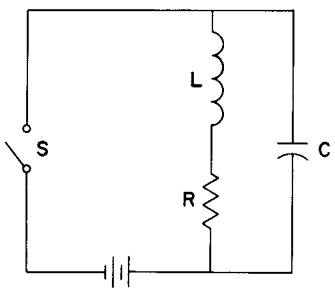


Figure 5. Simplified Equivalent of a Horizontal Deflection Circuit. The Power Tube is Essentially Switch S. Deflection Coils Are Simulated by the LRC Combination

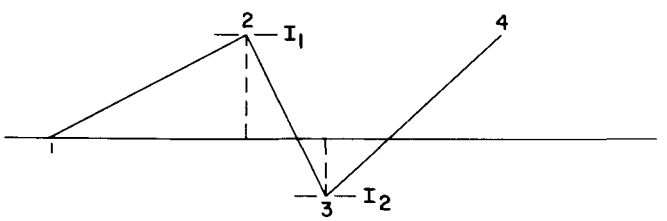


Figure 6. Horizontal Deflection Circuit. Output Current Waveform Showing Rise of Current to I_1 when Switch Is Closed. At Point 2, Switch Is Opened. Stored Energy Is Released Resulting in Oscillation of the LRC Circuit. After One-Half of a Cycle of Natural Frequency Operation (point 3) the Switch Is Closed and the Process Repeated

In terms of beam deflection, point 1 occurs when the beam is at the center of the screen, point 2 occurs at the right-hand edge of the screen, and point 3 occurs at the left-hand edge of the screen. The distance from point 2 to point 3 is the retrace time. The only way in which this circuit can provide linear deflection current is to reduce the circuit resistance to zero. During the period when the switch is closed, a negative resistance is needed to supply a reverse voltage which will cancel the voltage drop across the circuit resistance and, thus, insure linearity.

A damper diode is used to accomplish this function (refer to Fig. 7). Through the use of this diode, a boosted B-supply is created, which can supply limited power of a higher than B+ voltage external to the circuit increase. During retrace, the change of current from a positive peak to a negative peak results in a large back electromotive force (about 5000 volts positive at the plate of the deflection tube). By the addition of turns to the transformer and by the use of a high-voltage rectifier (V_3), the high-voltage pulse can be rectified and the picture-tube-anode voltage developed.

The horizontal output tube for the horizontal-deflection system must handle the power required to supply: deflection coil losses, linearity-transformer losses, transformer losses, power used in boosting B-supply voltage, and power used to produce picture-tube ultor voltage as well as the power needed by the filament of the high-voltage rectifier tube.

The amount of power actually needed depends on the picture-tube deflection angle, the picture-tube ultor voltage, deflection-coil factors such as number of turns and length of magnetic flux lines, effective yoke length and coil diameter, and the drain of the B-voltage-boost supply.

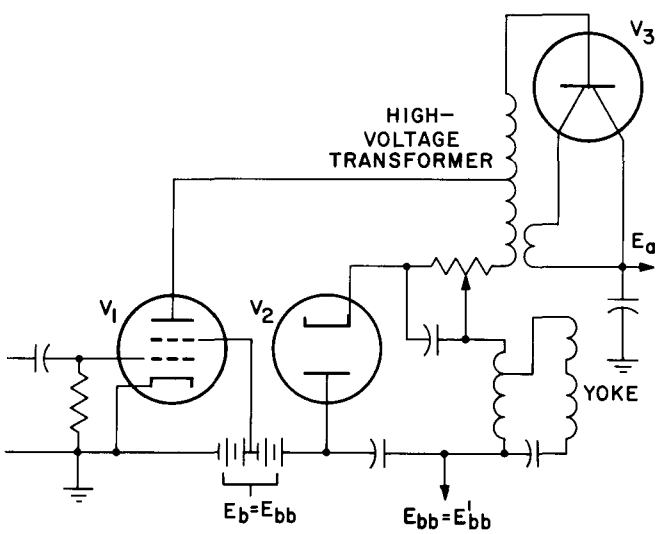


Figure 7. Simplified Horizontal-Deflection Circuit. V_1 Is the Horizontal Output Beam Power Pentode. V_2 Is the Damper Diode. V_3 Is the High-Voltage Rectifier which Rectifies the High-Voltage Pulses from the High End of the High-Voltage Turns and Supplies the Kinescope with its Anode Voltage. E'_{bb} is the Boosted Voltage Developed which Is Higher than the Supply Voltage E_{bb}

DESIGN EXAMPLE

A typical horizontal power circuit for a 17-inch, 110-degree picture tube using a 6DQ6-A horizontal-deflection beam power tube might have the following operating conditions:

B-Supply Volts	260
Ultor Voltage—Kilovolts	16.5 at 0 beam current 15.7 at 100 micro-amperes beam current
Cathode Current—millimaperes	130
Grid-No. 2 Quiescent Current—milliamperes	13.7
Grid-No. 2 Quiescent Voltage—volts	182
Grid-No. 1 Signal Voltage—peak-to-peak volts	135
Deflection Peak Plate Voltage—volts	630

In normal deflection circuitry, the ratio of the peak plate current of the horizontal-deflection output tube to the average plate current is approximately three. Therefore, peak plate current in this application is approximately (130 - 13.7) 3 or 350 milliamperes. Although analytical methods for determining the plate dissipation exist, they involve difficult measurements. As a result, the accepted method is to determine the tube bulb temperature during actual operation and then, keeping heater, screen-grid dissipation, and ambient conditions the same during the test, to set up a direct-current circuit and plot bulb temperature vs. different amounts of plate dissipation. In this example, plate power equals:

$$P_p = 13 \text{ watts}$$

$$P_{c2} = 13.7 \times 0.182 = 2.5 \text{ watts.}$$

From the high current required at zero grid-bias voltage in the above example, it is apparent that any triode design for this application would need a very large cathode and, therefore, that such a design would be impractical. Another limitation to the use of triodes arises from the difficulty of cutting triodes off during the retrace portion of the cycle. Pentodes and beam power tubes must be used.

A major part of the plate dissipation loss is due to the minimum voltage at the knee of the tube curve multiplied by the average plate current. (See Fig. 8.) More efficient operation could be obtained if the knee voltage were lower because the tube would then have less plate dissipation and more output. The screen-grid circuit is also quite critical. To obtain better stability of operation, lower knee voltage, and lower grid dissipation, it is usual practice not to obtain the screen-grid voltage direct from the B-voltage supply, but to use a lower value of voltage obtained through a screen-dropping resistor, properly bypassed. This method provides a lower knee voltage (and current), a lower screen dissipation, and a screen current having improved regulation. The resistance of the resistor is determined by the desired current level and the screen-grid dissipation rating of the tube. Screen-grid dissipation is normally the factor which limits the output of the deflection tube, in that it is usually the first tube rating to be exceeded. The second tube limitation in systems using a B-voltage supply of 250 volts or more is the peak positive-pulse plate voltage. In systems using lower B-voltage supplies the peak positive-pulse plate voltage attained is usually lower and, therefore, not limiting, but the cathode current increases.

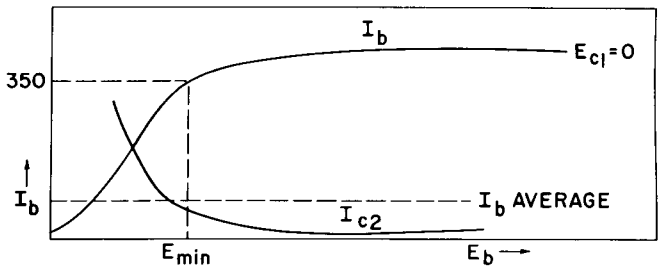


Figure 8. Zero-Bias Characteristics of the Horizontal-Output Beam Power Pentode. Energy which Is Dissipated in the Plate of the Tube Is I_b Average x E_{min}

DESIGN CONSIDERATIONS

Because high plate currents are required at the lowest possible screen-grid dissipation, beam power tubes of low-mu-factor design are used. A low-mu-factor design is used so that the tube will operate at the lowest possible screen-grid voltage which will give the necessary high plate current at the knee of the zero-grid-bias-voltage curve; a beam power tube is used to obtain a high ratio of average plate current to average screen-grid current.

In addition to normal beam power tube design criteria, the following points should be noted:

First, the compromise between a high current ratio and a low knee voltage should be carefully evaluated. With typical compromises, the knee voltage occurs between 30 and 40 per cent of the applied screen-grid voltage. That is, if the screen-grid voltage is 150 volts, the knee voltage of the tube would occur between 45 and 60 volts; however, the current ratio of a tube design giving a low knee voltage may not be as high as would be obtained from the design with a higher voltage knee.

Second, the tube design which has the lowest mu-factor should be attempted first. If the cutoff voltage desired is known, the new design should be adjusted to meet this requirement fairly closely.

Usually, the desired peak plate current on the zero-bias voltage curve is known. A design with a low mu-factor will give this current at a low screen-grid voltage. With proper suppression, the design should give

a low knee voltage. A design with a higher mu-factor will need more screen-grid voltage to give the desired plate current on the zero-bias-voltage curve and, thus will give a higher knee voltage.

Third, the design of the metal parts of the tube such as the beam-confining plate and the plate, should take into consideration the peak positive-pulse voltages that will exist on the plate. The distance between the plate and the beam-confining plate should be chosen to keep the voltage gradients between these parts as small as possible. In addition, it may be desirable to round the edges of these parts to avoid high voltage gradients at the edges. Any mica surface on which the plate bears should be kept at the greatest distance possible from other elements bearing on this mica and should be protected by a leakage slot in the mica. The spacing between the edges of the plate and the best means for visual alignment of the grids.

Fourth, the mica should be designed to obtain the lowest voltage gradient across its surface from the plate to the other elements. Slots should be used in the mica around the plate holes, the beam-confining plate holes, and other areas in high-voltage fields. Such fields can best be found by means of an equipotential plot. In adding slots, however, it is necessary not to weaken the mica by making the mica bridges too small.

Fifth, careful thought should be given to the use of the best available mica spray. The spray must not only limit leakage at low voltage between elements, but must also be resistant to charging and leakage effects at high voltage. In addition, the coating must have excellent adherence so as not to leave any bare spots on the mica. It is good practice to review the standardizing notices for the best available spray.

Sixth, the design of all parts should be such as to afford easy degassing of the parts since current in the grid circuit due to gas may cause the tube to exceed its dissipation ratings and run away.

Seventh, grid emission from either the control grid or screen grid can seriously reduce the output of the tube. Therefore, the grid temperature should be kept low: below 300 C for the control grid, below 450 C for the screen grid. Screen-grid emission can seriously impair the high-voltage performances of the tube. When the control-grid signal cuts off the tube, the plate voltage jumps to a value of many thousand volts. The voltage gradient at the screen grid then becomes high and causes the grid to emit the plate. If emission occurs, the back electromotive force in the transformer is reduced which, in turn, reduces the picture-tube ultor voltage.

Eighth, plate emission, during that portion of the cycle when the plate voltage may be below zero volts (after retrace) reduces the picture-tube ultor voltage, The tube designer, therefore, should keep the plate temperature low by using plate materials with good radiating properties and low gas content and, if necessary, radiating fins.

Ninth, as soon as a design has been established and the developmental samples have performed in a satisfactory manner, the tubes should be read for normal cutoff. These readings are usually taken at the same conditions which were used to check the lineup of the tube. Appropriate conditions are plate and screen-grid voltages of about 100 to 200 volts and the proper grid-bias voltage to give approximately one milliampere of plate current. The cutoff value obtained should be recorded in the test specification, because it is important that the cutoff value be maintained consistently in production.

Tenth, horizontal-deflection output tubes are especially critical to microphonism because the resultant electrical disturbances in the tube output degrades the linearity and causes visual distortion in the picture raster. Therefore, a tight testing control is needed to weed out production product which shows microphonism. In addition, it is important that the design engineer provide a structure which ties down all elements as rigidly as possible.

Eleventh, a disturbance called "snivets" is particularly annoying in the horizontal circuit. Snivets are vertical black bars, lines, or patterns which show up on the right-hand side of a kinescope picture. The effect seems to be caused by an instability in the virtual-cathode region when the anode voltage of the horizontal-deflection output tube is very low (see Fig. 9). It can

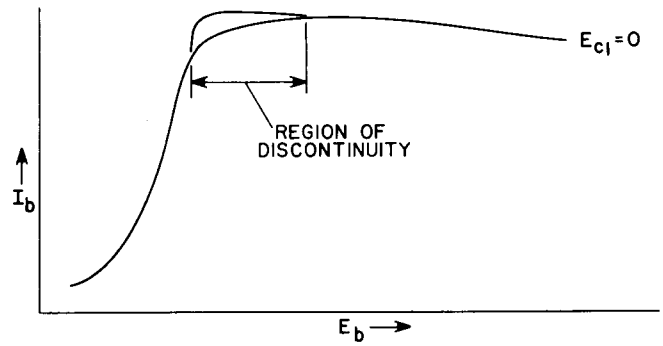


Figure 9. Horizontal Output Beam Power Pentode Plate Curve Showing Discontinuity of Plate Current at the Knee of the 0 Bias Trace which Causes Snivets

result in discontinuous jumps in current during the scanning cycle. These current changes are rapid enough to produce a frequency spectrum extending through the vary-high-frequency (vhf) band into the ultra-high-frequency (uhf) band. When aperiodic coupling is introduced between the tube and the tuner, the snivet signals are approximately the same in all uhf channels. To avoid this condition, the tube designer must take steps to provide proper suppression at the desired current level. Suitable tube design presents a major paradox to tube engineers. The better the design, that is, the sharper the knee, the higher the plate current at the lowest knee voltage, the more likelihood that snivets will be present. Conversely, the more rounded the knee and, thus, the less efficient the tube, the less the likelihood of snivets, but these methods usually involve additional parts or reduce output. The design methods, which have been tried with limited success, involve

varying the spacing between plate and screen grid by introducing either embosses or a metal part at right angles to the plate structure as shown in Fig. 10. The purpose is to make the plate surface irregular and, in this manner, to improve the stability in the virtual-cathode region.

Twelfth, in some designs, there is the possibility of high-voltage leakage due to the material deposited by the getter flash. If, for example, the getter is flashed in the dome of a tube, and the tube has a top-cap plate connection, then, the getter material becomes a part of the plate circuit. If leakage occurs between the getter material and any other electrode, the effect is an undesirable additional plate loading. If the leakage is sporadic, it can also, through radiation, cause a horizontal-synchronization disturbance in the horizontal synchronizing signal. The getter flash should be so localized that it does not provide a leakage path to any other electrode.

Thirteenth, if the getter is flashed in the stem, care must be exercised to: avoid depositing getter material on the mica where it could cause leakage across the mica surface; avoid the possibility of having the getter material form conductive paths between stem leads; leave adequate room for a decent-sized deposit of flash material because the gettering action depends upon the amount of getter surface.

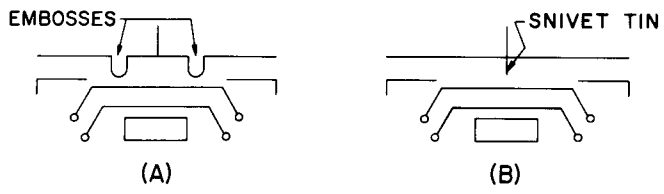


Figure 10. Two Methods Commonly Used With Some Success to Obtain Compromise between Snivets in the 0 Bias Trace of Beam Power Pentode and a Sharp 0 Bias Knee: (A) use of plate embosses to present an irregular plate surface; (B) use of a snivet tin for same purpose

Fourteenth, getters which are flashed against the wall of the tube opposite the plate raise the plate temperature and, as a result, should be used only with extreme care.

Black-and-White Picture Tubes

C. W. Thierfelder, et al*

Marion

A picture tube is a sophisticated cathode-ray tube used to convert intelligence from an electrical form into light. The conversion is accomplished by means of an electron beam which is focused to a small cross section on a fluorescent surface and is varied in position and intensity to reproduce a picture.

A picture tube generally consists of a bulb, fluorescent screen, aluminum backing and an electron gun. The bulb is a glass envelope which supports the screen and maintains a high vacuum. The fluorescent screen converts the electron-beam energy into light. Each excited phosphor crystal emits light uniformly in all directions. The light emitted toward the gun end of the tube is reflected forward by the aluminum backing. As a result, the light output is approximately doubled. Without the aluminum backing, the backward light would be diffused by multiple reflection from the inner bulb surface and would produce a general background illumination. The electron gun forms the free, thermionically emitted electrons into a beam and focuses the beam into a small spot on the fluorescent screen. The intensity of the beam is modulated by the video signal, and the beam is deflected to scan the screen in order to reproduce the picture.

At present, cathode-ray tubes are most widely used in black and white television service. In this application, the picture tube, or kinescope, is required to operate in a variety of receivers which have the following voltages available:

1. Ultor (collector and connected elements) voltage = 10 to 25 kilovolts
2. Accelerating electrode voltage = 200 to 500 volts
3. Focusing electrode voltage = -50 to +450 volts with no current flow
4. Video signal = 50 to 120 volts

Except for special kinescope applications, these voltage ranges are typical for kinescopes of any deflection angle.

The kinescope design must meet the standards of commercial picture performance and provide reliable service under these operating conditions. In general, the picture quality is determined broadly by the large-area and detailed contrast, focus, and light-output characteristics.[1] Two of these, large area contrast and light output, are significantly improved by a metal-backed (aluminized) screen. Detailed contrast is a function of the beam spot size. However, some improvement in detailed contrast may be obtained by aluminizing a screen, since aluminizing reduces the beam power requirements. Present aluminized kinescope designs provide large-area contrast ratios of over 100. A light output of about 30 foot-lamberts is obtained for an input power of 6.1 milliwatts per square inch.

Because of the many combinations of electrode voltages and beam currents possible, and since complete dynamic focusing is not feasible, an Einzel lens (a soft-focusing lens system) is used in current RCA designs. Einzel lenses are designed to be insensitive to fluctuations in electrode voltages. In general, acceptable picture focus is attainable by connecting the focusing electrode to any potential between 0 and +500 volts.

The beam current is proportional to the drive or modulating voltage applied between grid No. 1 and the cathode. Control of the cutoff range must be maintained since the cutoff voltage of the kinescope limits the amount of effective drive voltage which can be used. Because the amount of drive voltage generated by the television receiver is not a function of kinescope cutoff voltage, the electron gun must be designed so that:

1. A substantial portion of the drive range of the kinescope be used to provide about 6000 tone values, which can best be described as different light levels from black to peak brightness

2. The cutoff voltage must be high enough to accept sufficient drive voltage for the cathode currents needed to achieve the required peak highlights

ELECTRON GUN DESIGN CONSIDERATIONS

DESCRIPTION AND OPERATION

An electron gun is a device which produces, controls, and accelerates free electrons into a well-defined beam. The gun may or may not include a focusing lens by which the electron beam may be converged on a given surface.

The gun contains a cathode, from which the electrons

*Authors: C. W. Thierfelder, D. C. Ballard, R. E. Benway, E. O. Hanson, R. J. Konrad, R. L. Leigh, R. E. Salveter, and L. E. Wuellner

are accelerated toward the ultor and controlled by the potential difference between the cathode and the control grid, known as grid No. 1. In a kinescope, these electrons may be accelerated to the desired velocity by any number of electrodes. The electrons are focused to converge upon the fluorescent screen in a small, symmetrical spot.

Magnetic-focus guns normally do not contain the means for focusing. The magnetic field required for focusing is usually formed by a component external to the kinescope. The gun, therefore, generally contains only beam-forming elements, the cathode, grids No. 1 and No. 2, and a final accelerating element known as grid No. 3.

On the other hand, electrostatic-focus guns contain electrodes which establish electrostatic focusing fields upon the application of the rated voltages. Electrostatic-focus guns contain beam-forming elements in addition to the focus electrodes.

In addition to the focusing media, many electron guns contain a tilted or offset electrostatic lens, called an ion trap, which requires a magnetic field component external to the kinescope for its operation. Proper spot centering or picture centering in a television receiver requires ion trap guns to be tilted with respect to the axis of the tube, thus the term "tilted gun." Because of this, the term "straight gun" has been limited in usage to describe only those electrostatic-focus guns which do not contain ion-trap lenses.

ELECTRON LENS THEORY[2, 3, 4, 5, 6]

Basically all vacuum tubes employ the effects of electrons moving under the influence of electric or magnetic fields. The electron dynamics principle represents the foundation on which the physical and analytical concepts are built. According to Gauss's law, an electric charge constitutes a source of electric intensity or electric flux. Electric flux is customarily represented by lines starting on a positive charge and terminating on a negative charge. The force F_{ES} on an electron moving through an electrostatic system in the direction s with the potential difference of E is given by:

$$F_{ES} = -eE_s$$

The magnetic force F_M on an electron is determined by the product of the velocity v of the electron and the component of the magnetic field B normal to the direction of the electron.

$$F_M = -evXB$$

The velocity of an electron or ion moving through a potential difference E is determined by:

$$v = \sqrt{2E\,\frac{e}{m}}$$

Because the velocity is dependent on mass, the force is dependent on mass in magnetic lenses; whereas the force in electrostatic lenses is independent of mass.

An example of an electrostatic-lens system is electrostatic deflection produced by two deflecting electrodes. The deflection produced can be expressed by:

$$D = \frac{aLE_1}{2bE}$$

Where, E_1 = Potential difference between deflecting electrodes in volts
L = Distance from center of deflecting electrodes to screen in centimeters
a = Length of deflecting electrode in centimeters
E = Acceleration voltage applied to electrons at deflecting electrodes
b = Spacing between deflecting electrodes in centimeters
D = Deflection of beam at screen in centimeters.

Magnetic deflection can be represented in a general way by:

$$r = \frac{1}{B}\,\frac{2mE}{e}$$

Where, B = Magnetic field strength in gausses
m/e = Mass-to-charge ratio of electron
E = Accelerating voltage in volts
r = Radius of curvature in centimeters

Practical applications of kinescopes require some method of causing the beam to scan the screen surface.

DEFLECTION

For commercial television, scanning provides a rectangular raster with about a 3:4 aspect ratio.

Magnetic deflection, the method of scanning commonly used, makes use of coils placed adjacent to the neck of the tube in the reference-line region (the approximate intersection of the neck and the funnel of the kinescope) to deflect the beam. Two pairs of coils are used: one for horizontal deflection, and the other for vertical deflection.

The deflection of an electron traveling at a constant velocity requires no power. However, power is required to supply the electrical losses in the magnetic deflection coils used to establish the magnetic fields. This is a major disadvantage of magnetic deflection. This limitation is, however, outweighed in practical applications by the advantage of being able to place the deflection components outside the picture tube so that the kinescope neck itself can be made small. Coils provide a further advantage in that they may economically be shaped to give high deflection sensitivity and to deflect high energy beams to wide angles. Another advantage of the use of deflecting coils is that they become part of the scanning circuit so that the high voltage for acceleration of the beam electrons is obtained as a by-product.

Electrostatic deflection, on the other hand, has the advantage of requiring little power since the deflection is a result of electrostatic charges on deflecting electrodes within the kinescope. The voltage producing

this charge, however, is required to be an appreciable fraction of the beam-accelerating voltage of the electrons at the time of deflection and, therefore, causes insulation problems. Accurately obtaining and controlling the high deflecting voltage at the scanning rate is also a problem. Electrostatic deflection is generally limited to small deflection angles (below 40 degrees) due to deflection distortion caused by the restricted field size and the fringe-field aberrations. Additional information on this topic will be found in the article on "Oscillograph Tubes."

With magnetic fields, deflection defocusing occurs because the beam has a finite diameter at the time of deflection. When the beam is deflected off the axis of the tube, there is a difference in the magnetic field intensity on opposite sides of the beam which produces deflection distortion by deflecting the sides of the beam to different amounts. The effect is to provide additional focusing of the beam in the direction of scanning and to distort the shape of the spot.

A somewhat different effect is observed in electrostatic deflection. The distance between the nearest deflecting electrode and the closer edge of the beam is the variable which results in greater deflection of the near electrons. The resultant spot is therefore underfocused in the direction of scan and is thus distorted.

This deflection defocusing phenomenon adds to the difficulty of gun design for wide-deflection-angle tubes in that the electron gun must produce a small beam diameter in the deflection region.

The tube designer must work closely with the designers of deflection yokes and circuits to obtain optimum bulb and yoke contours.[7] Considerations must include beam diameter, neck length, neck diameter, gun length, methods of beam centering, the requirements of deflection sensitivity, and the reference line contour.

Modern television receivers, with their compact cabinets, require very short picture tubes. The requirement for short neck lengths results in short guns, which are immersed in the deflection field. The resulting problems require very close liaison between the tube designer and designers of sets and components. The highly aberrated deflection field behind the yoke may be positioned behind the cathode of ultra-short-neck kinescopes, while the main field is used to deflect the beam before, during, and after its passage through the focusing lens. The beam passes through the high-aberration sections of the focus lens with consequent distortion of the focused spot at the outer edges of the picture. This effect can be decreased by minimizing the beam diameter in the focus lens region.

The wide deflection angles used for modern television often require the use of small magnets in conjunction with the deflection yoke to correct the raster shape. These magnets also add distortions to the beam. The tube designer must, therefore, obtain small beam diameters throughout the deflection region, and a smaller spot size at the screen than otherwise would be required so that the effects of these distortions are less noticeable.

To compensate for these distortions, dynamic focusing, which is the application of an ac focusing voltage at synchronized horizontal scanning frequency, is occasionally used with electrostatic focus tubes.

Picture centering is usually obtained with a magnet of variable strength and direction located on the tube neck immediately behind the deflecting yoke. The distortions that result from this beam bending are similar to the effects caused by deflection before focusing. The centering correction requirement, which is determined by the kinescope spot landing, horizontal-deflection phasing, scanning linearity, and extraneous magnetic fields, must be kept to a minimum. The effect of the vertical component of the earth's magnetic field, being relatively constant throughout the continental United States, is best corrected through the use of a fixed permanent magnet such as a magnetized yoke band. This method allows the correction to be made near the apparent center of deflection and minimizes the beam distortions caused by asymmetrical focusing and deflecting fields.

TRIODE AND TETRODE CHARACTERISTICS

The source of electrons for picture tube guns is an indirectly heated cathode, consisting of a nickel-alloy sleeve with a closed end on which the cathode coating is sprayed. Grid No. 1 is generally made from a closed-end metallic cylinder with a centrally located aperture (between 0.022 and 0.036 inches in diameter) formed in the closed end. The cathode is mounted within this cylinder with the cathode coating spaced approximately 0.0035 - 0.0075 inch from the inside end of the aperture. Negative voltage applied to grid No. 1, or a positive voltage applied to the cathode, controls the beam current.

Coaxial with grid No. 1 and spaced along the axis is another cylinder that operates at a positive potential and provides a field to accelerate the emitted electrons. For a triode type of gun (see Fig. 1), this electrode is operated at the full accelerating potential.

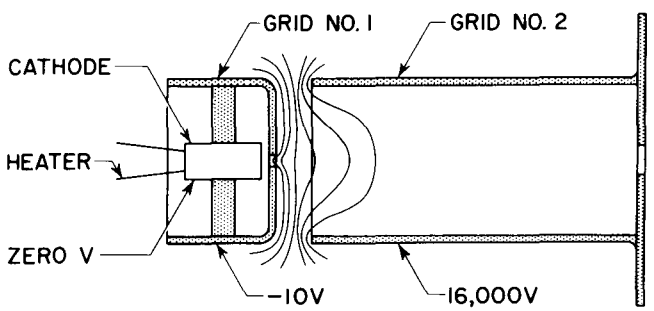


Figure 1. Magnetic-Focus, Triode-Type Gun

Any changes in accelerating voltage affect the beam current, cutoff voltage, and image reproduction. These effects are minimized by the addition of grid No. 2, the screen grid, to the conventional triode type of gun. Grid No. 2 is placed between grid No. 1 and the ultor or grid No. 3 (see Fig. 2).

Grid No. 2 operates at 200 to 500 volts and acts with

the control grid to form the first electron lens, the cross-over forming lens, and acts with the grid No. 3 to form the prefocusing lens. More important, it acts as a shield to minimize the effects of variation of grid No. 3 voltage on the cathode current, cutoff voltage, and image reproduction. Therefore, picture tubes using a tetrode gun may be operated at different final accelerating voltages, depending on the set manufacturer's design, without affecting the modulation characteristics.

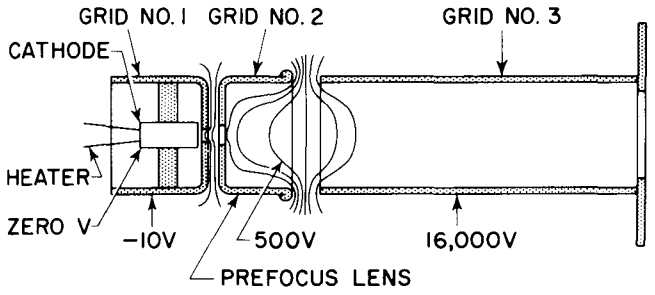


Figure 2. Magnetic-Focus, Tetrode-Type Gun

CUTOFF VOLTAGE

The cutoff voltage is the lowest negative voltage which, when applied to the control grid with the cathode grounded, will reduce the beam current to zero. A less negative voltage, or positive video signal, applied to the control grid increases the beam current and the light output as illustrated in Fig. 3.

The similarity of the curves in A and B of Fig. 3 indicates that the light output is directly proportional to the beam current over a wide range. Furthermore, the zero-bias, or maximum, beam current obtainable without driving the control grid positive with respect to the cathode varies with the cutoff voltage according to the following relationship:

$$MI_b = K \phi V_{co}^{1.5}$$

where MI_b is the maximum beam current, K is a constant determined by the geometry of the gun (generally equal to 3), ϕ is the cathode quality factor (which has a nominal value of 1), and V_{co} is the cutoff voltage. Since light output is directly proportional to MI_b, examination of the equation will point out the importance of the cutoff voltage in determining the peak highlight brightness obtainable from a picture tube.

Design Parameters

The design parameters affecting cutoff voltage in kinescopes are related by the following general formula. The exact relationship for a particular grid-cathode geometric configuration, though similar, would have to be determined and verified experimentally:

$$V_{co} = K \frac{E_{c2}\, D_1^3}{f(a, b, t, D_2)}$$

Where V_{co} = grid-No. 1 cutoff voltage
D_1 = grid-No. 1 aperture diameter
D_2 = grid-No. 2 aperture diameter
a = grid-No. 1-to-cathode spacing
b = grid-No. 1-to-grid-No. 2 spacing
t = aperture thickness of grid No. 1
E_{c2} = grid-No. 2 voltage
K = a constant of proportionality

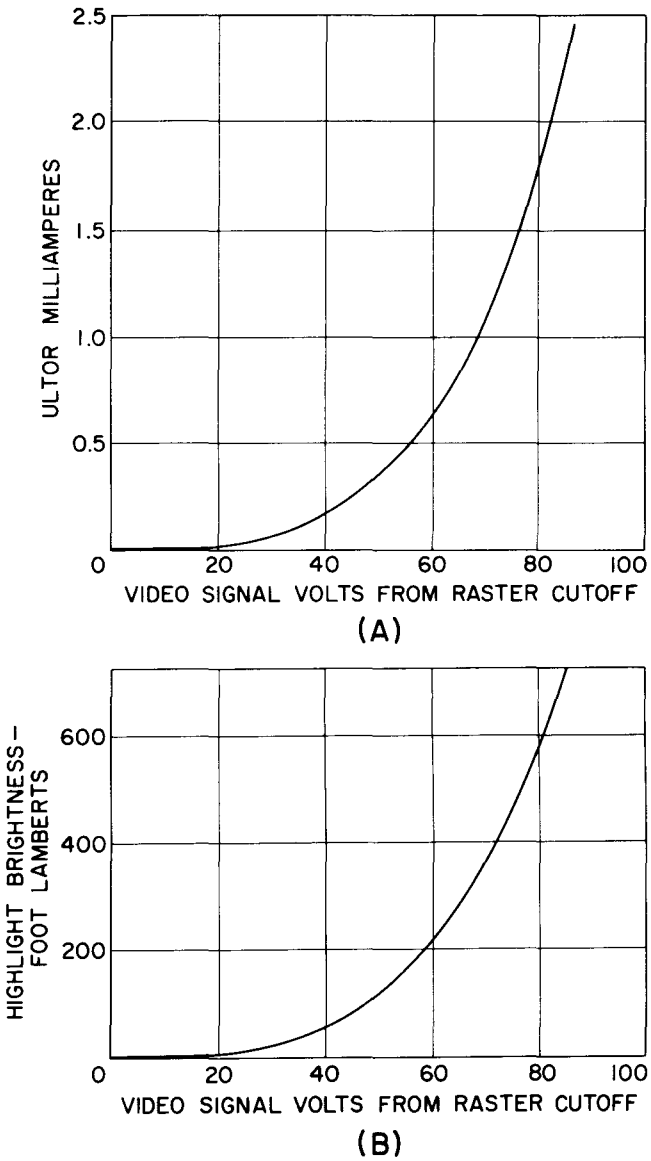


Figure 3. Grid Drive Characteristic of RCA-23MP4

In practical applications involving conventional kinescope grid-cathode constructions, the relationship can be approximated by:

$$V_{co} = K \frac{E_{c2}\, D_1^3}{a, b, t}$$

The value of K can be considered constant provided that changes in the parameters are small. D_2 is omitted

because it has little effect for small changes and is normally held constant as other factors are varied.

In 1953 a series of curves (Fig. 4) were prepared by H. C. Moodey which are useful in determining the approximate change in cutoff voltage caused by changes in one or more of the parameters given in the general formula. Origin, reference material, and notes concerning these curves follow:

Curve for	Sources of Data[4, 8, 9]
D_1	Moss indicates that this third-power relationship becomes a 2.5 power relationship for unusually small values of D_1 and for very large values of a. Allard shows 2.5 power; Maloff and Epstein show 2.0 power for medium values, increasing to over 4 for high values of D_1.
D_2	Moss indicates D_2 is independent of V_{co}.
a	Moodey — large quantity of data: ends of curve are dashed because "a" vs. V_{co} is somewhat affected by other factors. For small changes in "a" and conventional gun dimensions, $V_{co} = ka^{-n}$ in which: n = 1.0 (source: Moss) n = 0.44 (source: Moodey)
b	Moss
E_{c2}	Moss and Allard, substantiated by extensive RCA data.

Additional Factors

Aperture thickness of grid No. 1, t: Moss indicates V_{co} varies inversely as t (or $t^{0.5}$ for very large "a"), and Allard shows V_{co} varies inversely as the 1.25 power of ("a" plus t).

Mechanical Variations

There are many mechanical variations in the manufacturing and assembly processes which cause variation in the design parameters. The most significant mechanical variations are: tolerances allowed for fabrication and wear of perforating tools; variations in incoming material; thinning of materials during drawing operations; variations in coining operations designed to control thicknesses of very thin apertures; variations in camber of the aperture surfaces; variations in height of the grid No. 1-to-cathode spacer; variation in distances from the top of the cathode to the ceramic; variations in cathode coating thickness and density; variations due to expansion or distortion of parts upon heating during processing and during normal operation; movement of the cathode sleeve within the ceramic; change due to fusion of the cathode coating; and mechanical variations introduced in assembly techniques.

From a practical standpoint, more than 50 per cent of the variation in cutoff voltage may be directly attributed to variations in the spacing between grid No. 1 and the cathode. This high percentage is due to the many mechanical variations affecting this parameter.

Practical Manufacturing Techniques

Measuring, sorting, and matching subassemblies for each grid assembly is the most common method of fixing and controlling the grid No. 1 to cathode spacing. Sometimes, subassemblies are pre-sized before measuring and sorting to reduce the number of categories that require matching. Examples of these sizing operations are, sizing grid No. 1-to-cathode spacers or using a necked down grid No. 1 cup with the necked down section presized, and sizing of the cathode-cap height. Measuring the distance from the bottom of the grid No. 1 aperture surface after positioning the spacer in the grid No. 1, prior to the selection of a matched cathode assembly, will minimize the effect of the grid No. 1 camber on the grid No. 1-to-cathode spacing variation. Another important factor in controlling grid No. 1-to-cathode spacing is the control of the thickness of the sprayed cathode coating and its density.

A relatively new process for controlling cutoff voltage utilizes a capacitance-comparison-bridge circuit which will measure very small values of capacitance. By setting up this bridge circuit to balance the cathode-to-grid No. 2 capacitance of a mount that has a bogie cutoff voltage, other mounts may be assembled using random parts instead of selected and matched parts. After grid No. 2-to-cathode spacing has been adjusted to balance the bridge, the assembly can be welded or crimped in place. Special circuits have been designed in conjunction with this capacitance-comparison bridge which balance out grid No. 1-to-cathode and grid-No. 1-to-grid-No. 2 capacitance. The No. 1 grid becomes part of a shield to prevent stray capacitance from affecting the adjustment. (See Fig. 5.) Effectively, this bridge measures a small capacitor illustrated in Fig. 5 where one plate is a small circle of area A1, slightly larger than the grid No. 1 aperture, and the other plate has an area A2-A3, where A3 is the diameter of the grid No. 2 aperture. In most picture-tube electron guns, this capacitance is less than 0.004 $\mu\mu$f.

Practical techniques for measuring and inspecting grid No. 1-to-cathode spacing after assembly are made possible by use of the following types of tools: mechanical contact or dial indicators, calibrated microscope; measurement of voltage breakdown, air gage, capacitance-comparison bridge, and optical comparator.

PREFOCUSING LENS

The electron emission from the cathode is usually formed, by the electron lens field adjacent to the cathode, into a narrow beam having a minimum cross-sectional area near the cathode surface. This minimum cross section of the beam is known as the first crossover. The size of the crossover is determined by the distribution in velocity and angle of the emitted electrons as they strike the focal plane. As the beam passes through this crossover region, it diverges so that subsequent lens fields are required to cause the beam to converge back to the second point of minimum cross-sectional area at the fluorescent screen of the tube.

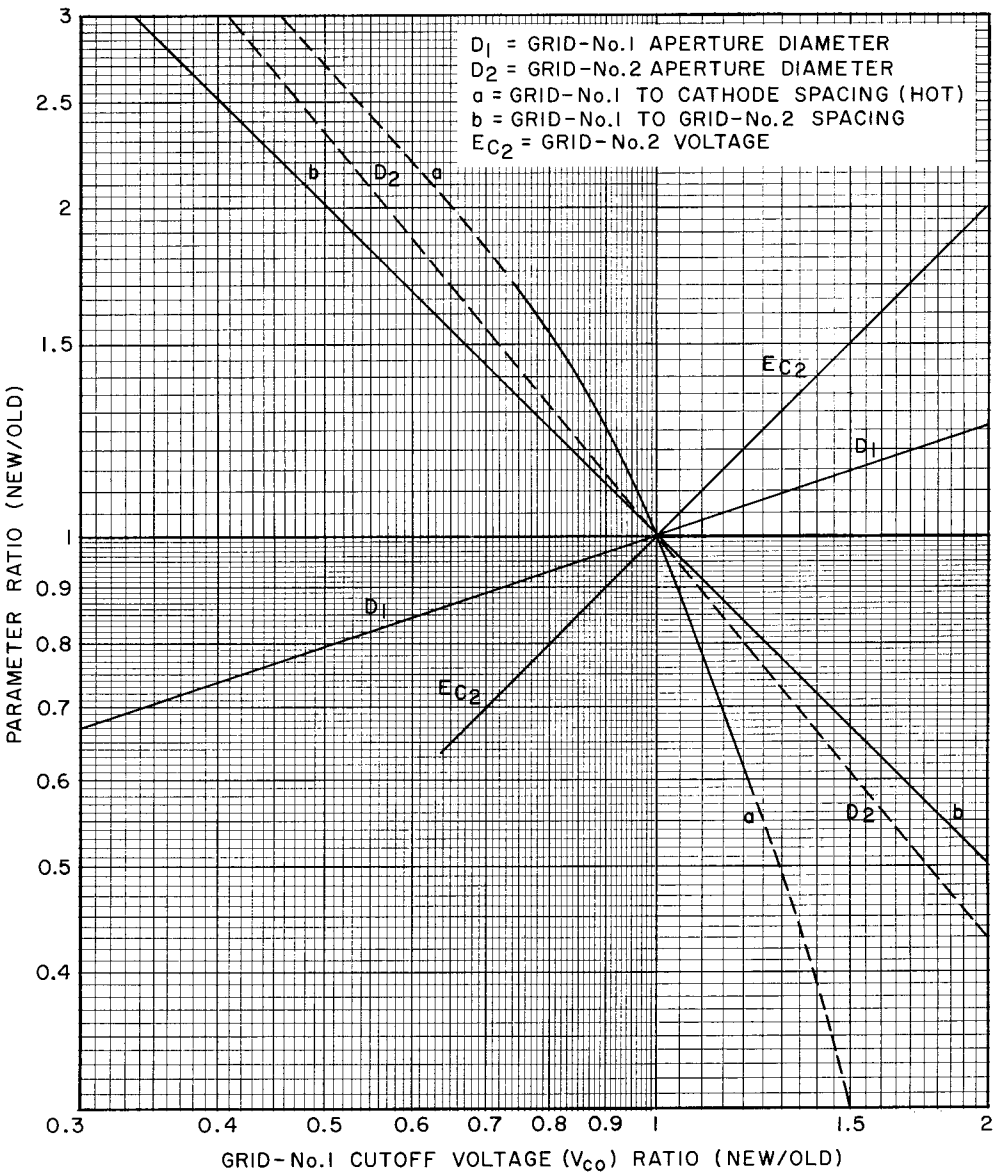


Figure 4. Approximate Variation of C_oE_{c1} with Change in Design Parameters

The electron beam may be focused by at least two electron-lens fields: a prefocusing field and a principal focusing field (see Fig. 6).

The first crossover of the beam serves as the image for the focusing lenses and forms a focused spot on the screen. The prefocusing-lens field reduces the size of the electron beam as it passes through the principal focusing field, thus reducing the aberration effects of the focusing lens by confining the beam to the more uniform center portion of the lens.

To provide the optimum prefocusing-lens field for a gun structure, the ratio of the diameter of the grid No. 3 aperture to the grid No. 2 aperture should be between 5:1 and 1:1, and the potential distribution between the two accelerating electrodes should be about 5 kilovolts per 0.040 inch of separation. This optimized prefocusing lens design permits the lens field produced between grid No. 2 and grid No. 3 to penetrate the grid No. 2 aperture and produce a strong converging lens effect on the electron beam in the crossover region. The field penetration that results is sufficient to shape the field between grid No. 1 and grid No. 2 in such a manner that a strong converging lens field is provided between the crossover and the cathode coating. This converging lens provides a smaller and better defined crossover region especially under high drive conditions, where the space-charge effects are reduced by accelerating the electrons to a higher velocity. This converging-lens field also reduces the divergent angle of the electron beam as it passes out of the crossover region. The reduction of the divergence of the electron beam passing through the crossover region results in a smaller beam size. The beam is now more confined to the center of the focusing lens, and less of the beam passes through the outer portions of the lens field, where the aberration effects become excessive. A smaller beam diameter in the electron gun also results in a smaller beam diameter in the deflection region of the yoke and thus significantly reduces deflection distortion and defocusing.

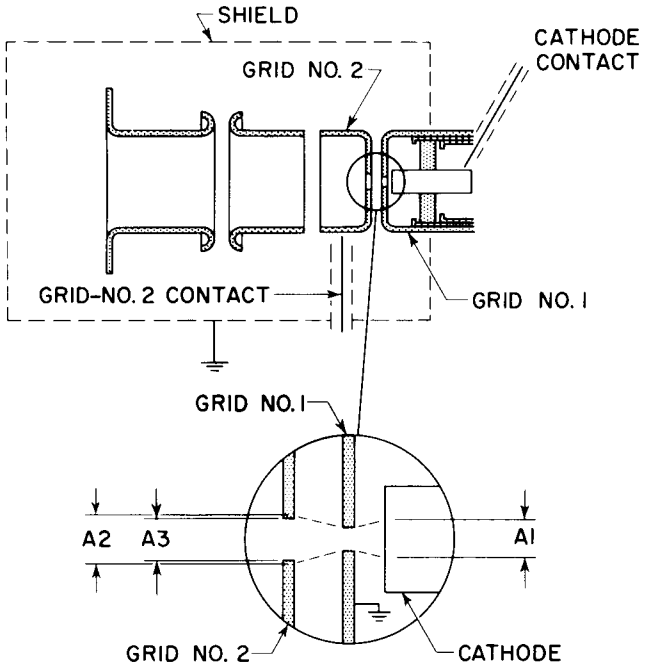


Figure 5. Capacitance in Picture Tube Electron Gun

ION TRAP LENS

Since electrons and singly charged ions have equal charges, they are deflected identically by electrostatic fields. The tilted lens (formed by grid No. 2 and grid No. 3) of RCA guns bends both the electron and the negative-ion beams along the same path in the direction of the tilt. The strength of these lenses, or the amount the beam is bent, is determined in general by the difference in potential between grid No. 2 and grid No. 3. If this difference is zero, the beam is not bent and electron and ion paths are nearly straight-line continuations of the paths through the grid No. 1 and grid No. 2 (Fig. 7A). If the potential difference between grid No. 2 and grid No. 3 is great enough, the paths are bent so that the beams do not emerge through the grid No. 3 (masking) apertures and are collected (or trapped) on the electrode (Fig. 7B). Similarly, positive ions, formed within grid-No. 3 cylinder will not pass through the apertures of grid No. 1 and grid No. 2 and are collected on grid No. 2.

The limiting trapping voltage is an important consideration in an ion trap design. This voltage is defined as the minimum grid No. 3 voltage required to cause the beam to just not clear the grid No. 3 aperture when the grid No. 2 voltage is at its maximum rating. The limiting trapping voltage is therefore the minimum ultor voltage at which a nonaluminized kinescope can be operated during its life with absolute assurance that no ion spot will form on the screen. The angle through which a beam will be bent can be extrapolated from data supplied by Moss, et al.[10]

If a magnetic field is applied normal to these beams in the region of the lens formed by grid No. 2 and grid No. 3, it is possible to bend the electron beam back through the masking aperture while the ions remain trapped (Fig. 7C). The electron beam is deflected through a greater angle than the ion beam by the magnetic field because of the difference in mass between an ion and an electron.

If the magnetic field is not relatively uniform, the deflection of different segments of the electron beam will be nonuniform since the electron beam has a finite diameter. The result is spot distortion which appears as poor focus.

The magnetic field distributions of most permanent-magnet (PM) beam benders currently in use for renewal kinescopes are extremely nonuniform. However, most of the nonuniformity occurs in a region near the magnet. The field near the open end of the pole pieces is relatively uniform.

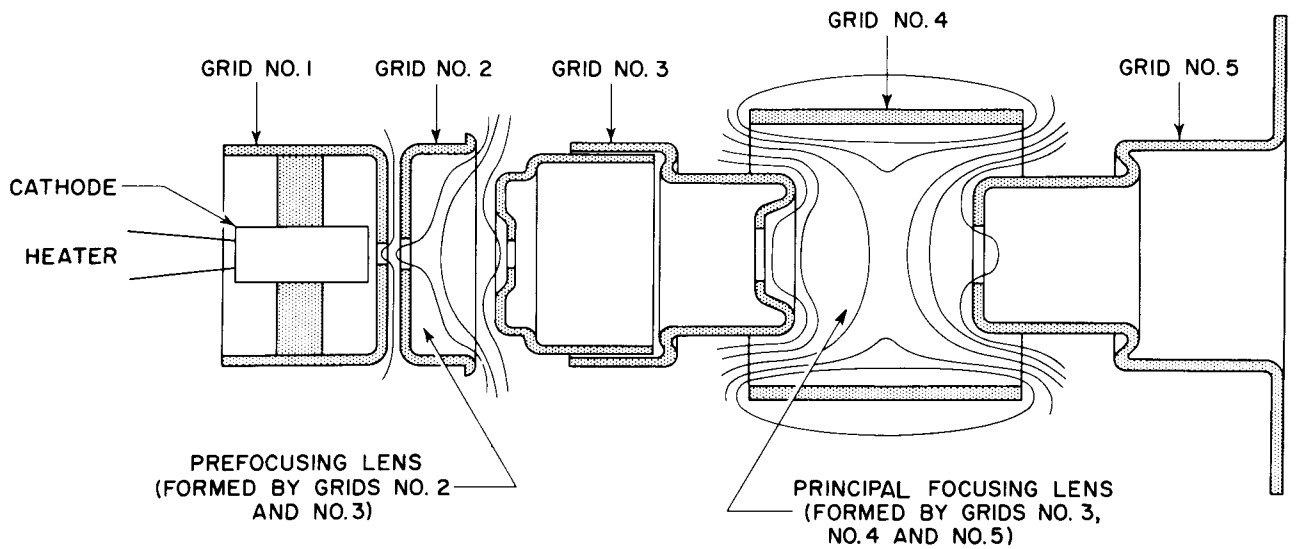


Figure 6. Electrostatic-Focus, Straight-Type Gun

A magnet which has a stronger field than that required to return the beam through the grid No. 3 aperture may be utilized by pulling it back toward the base away from the lens formed by grid No. 2 and grid No. 3. This adjustment causes some of the magnetic lines of force to pass behind the cathode, where they have no effect on the beam.

An ion-trap type of electron gun is tilted in the neck of the kinescope in some designs merely to cause the

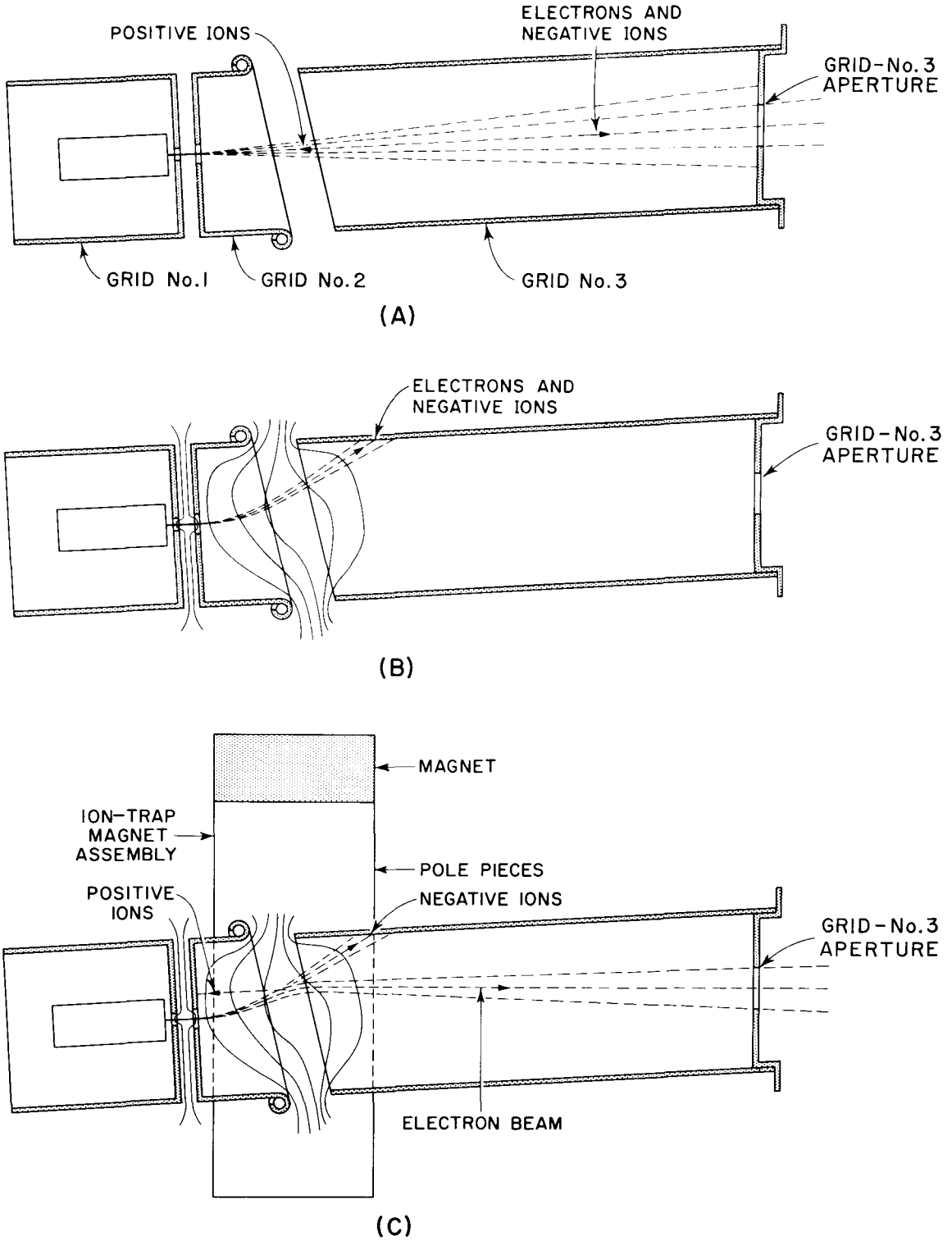


Figure 7. Approximate Paths of Electron and Ion Beams Through Mount: (A) grid-No. 3 voltage = grid-No. 2 voltage — no ion-trap magnet; (B) grid-No. 3 voltage>>grid-No. 2 voltage — no ion trap magnet (C) grid No. 3 voltage>>grid-No. 2 voltage — with ion-trap magnet.

electron beam to travel along the axis of the tube after it is deflected by the ion-trap-magnet field. The location of the ion-trap magnet axially along the tube neck has some bearing on spot centering because the bending of the beam takes place in the region of the magnet, and the beam location with respect to the tube axis varies with respect to distance. Since the change in centering is small, the angle at which the beam enters the focusing lens is very nearly independent of ion-trap strength. Consequently, the strength of the ion-trap magnet used has little effect on focus performance.

In general, ion-trap guns are no longer used in modern picture tubes. This is because of the superior focus characteristics obtainable with straight-gun designs, the trend toward short-neck, compact kinescope designs, and the use of aluminized screens, where the aluminum serves as a trap to prevent ion burning of the screen.

FOCUSING LENSES

Magnetic Lens[3]

The magnetic lenses that are commonly used for picture tubes are very short lenses produced by a gap in an external PM or electromagnet (EM) assembly. This type of focusing has become less popular in recent years due to the advent of compact short-neck kinescope designs. The magnetic lens also represents, to the set manufacturer, an extra component that he has to buy, mount, and adjust. This cost is eliminated when an electrostatic-focus tube is used. Aberrations that are present in magnetic lenses are very similar to those found in electrostatic lenses; however, since the magnet is located outside the neck where it has an advantageous ratio of lens diameter to beam diameter, a minimum of aberration is obtainable. Care in adjustment, to insure good rotational symmetry and accurate alignment of the lens with the axis of the beam, is required to avoid serious aberrations.

Electrostatic Acceleration Lens

In general, an electrostatic acceleration lens can be formed by placing a potential difference between two coaxial cylinders or apertured disks or a combination of both that have been electrically insulated from each other. If an electron beam is passed through such a system, the electrons in the beam converge to a point beyond the lens determined by the potential difference, geometric configuration of the lens, initial velocity of electrons entering the lens, and the final velocity of the beam electrons. The functioning of this system can be visualized by considering an electron that enters the system at an angle, other than perpendicular, to the equipotential lines of the field formed between the lens elements. The path of the electrons is bent toward the normal to the equipotential lines by the electrostatic field. The amount of bending is dependent upon the velocity of the electrons and the strength of the field perpendicular to the direction of electron motion. The shape of the field produced between the coaxial cylinders or apertures can be controlled by their geometry to form the desired electron lens.

To minimize lens aberration, nonrotational symmetry of gun parts must be avoided, and axial alignment of the gun parts must be maintained. To keep aberration to a minimum, it is considered good practice to use only a small section of a large lens for focusing.

Because of the high sensitivity of an electrostatic-acceleration lens, the voltage applied to the lens must be adjustable. A slight change in voltage will correct for slight geometrical changes due to tolerances in acceleration-type lenses. Because adjustment is possible, the tolerances on spacing and parts diameters of an electrostatic acceleration lens are not as critical as those of an Einzel lens.

Because electrostatic acceleration lenses are so sensitive to slight changes in voltage, the voltages applied must be regulated for some applications. Another disadvantage of this lens is that the first element sometimes requires a voltage source other than that used for the second element.

A low-voltage electrostatic acceleration lens has been developed by the choice of proper spacing and electrode geometry. The focusing voltage for the construction illustrated in Fig. 8 operates between 0 and 300 volts. The major disadvantage of this electrostatic acceleration lens is its sensitivity to changes of focal length with slight variations in grid-No. 2 voltage, ultor voltage, and ultor current.

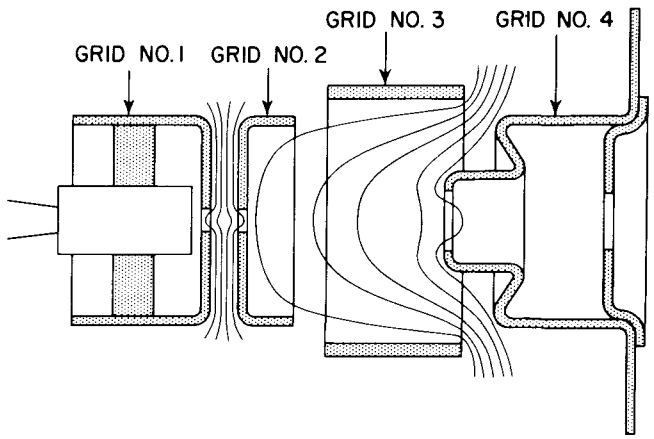


Figure 8. Low-Voltage, Electrostatic-Acceleration-Type Focusing Gun

The major advantages of the lens are: (1) performance is compatible with that of a compact gun for short-neck-length conditions under which the focus lens has to be located in the deflection field, (2) the high voltage stability is excellent, and (3) spherical and coma aberrations are much less than those of an Einzel type lens. When a slight gun misalignment occurs causing the beam and the focusing lens to be misaligned with respect to each other, the electrostatic acceleration lens produces less spot deterioration than an Einzel lens.

A prototype high-voltage, electrostatic acceleration lens is widely used for color kinescopes, projection kinescopes, and other special-purpose cathode-ray tubes. This type of lens is illustrated in Fig. 9.

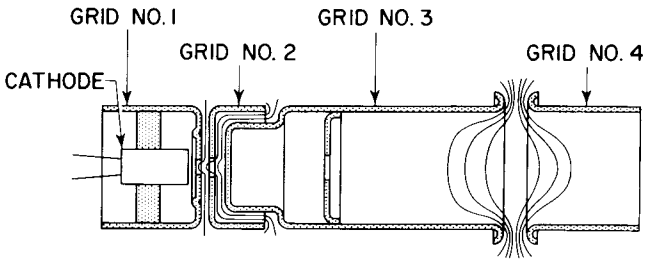


Figure 9. High-Voltage, Electrostatic-Acceleration-Type Focusing Gun

Major advantages of the high-voltage electrostatic acceleration lens are: (1) its spherical and coma aberrations are much smaller than those of other electrostatic lenses, (2) the high-voltage lens can also focus high beam current without excessive spot deterioration, and (3) the electron-optical spot magnification is readily controllable by design. The high-voltage stability is excellent and generally compatible with operation up to 100 kilovolts.

The disadvantages of the high-voltage lens are: (1) the focal length varies with variations in ultor voltage and current; (2) the additional circuit required to supply the focus voltage (approximately one-fifth of the ultor voltage) adds to the cost; and (3) a long neck is required.

Unipotential Electrostatic Lens[11]

The Einzel, or unipotential, lens* is one having the same refractive index on both sides of the lens. Equal refractive indices are obtained by connecting together the two outside elements of this three-part lens so that a field is produced between the center element and the outside elements. (See Fig. 10.)

The center element is tied to ground potential or some other desired voltage. Proper selection of electrode geometry and spacings will form a field that will focus the electrons on a prescribed plane some distance from the lens. Thus, this lens system uses the same voltage to accelerate the electrons and to focus the beam. The fact that only one voltage is required is an important economic advantage that has made this lens a very popular one, and is in part responsible for its wide use.

When this lens is adjusted geometrically to have the central, or focusing, electrode operate at or near ground potential the focal length will remain constant in spite of wide variations of the outer-electrode voltage. This characteristic permits the tube designer to optimize all other lenses for minimum sensitivity to voltage variations. Picture tube guns have been designed, and are widely used, in which the focus is virtually independent of the applied voltage and other operating conditions. This feature is extremely important to set designers, who have to provide for a wide variety of final voltages on different receiver models. It is equally important to the consumer because it insures optimum focus regardless of operating conditions. The elements of the unipotential lens can be formed by disks, cylinders, or apertured cups. The most popular form uses a cylinder for the central, or focusing electrode. A smaller apertured cylinder projects axially into both ends of the center element (see Fig. 10A). This configuration shields the lens from the effects of extraneous electrostatic fields. The spacing and electrode geometry are then adjusted so that the electron beam will focus on the screen when the desired low voltage is applied to the focusing electrode.

*Other names for the unipotential lens include: univoltage lens, saddle lens, and symmetrical lens.

LENS ERRORS[2, 12, 13]

Most electrostatic and magnetic fields with axial symmetry focus electron paths close to the axis and act much like a lens focusing a light ray. Unlike a light-focusing lens, an electric or magnetic field is continuous throughout the space. As a result, one has much less freedom in the design of electron lenses than one has in the design of light lenses. Because of this inherent lack of freedom, electron lenses have very severe aberrations. In analyzing these aberrations it is important to separate them into the most important component defects arising from different causes to consider each separately.

Aberrations of electron lenses include: (1) distortion, (2) astigmatism, (3) coma, (4) curvature of field, and (5) spherical aberration. The first four, because they affect only off-axis rays, are less significant for electron lenses used in picture tube guns than is spherical aberration. When rays coming into the lens nearly parallel to the axis, but at different radii, cross the axis at different distances beyond the lens, the effect is called spherical aberration. Both electrostatic and magnetic lenses have spherical aberration. Because of the smaller diameter of an electrostatic lens, however, the effect of this aberration is much more significant for electrostatic lenses than for magnetic. Spherical aberration is usually severe in short-focal-length lenses. The only practical way known to minimize the adverse effect of this aberration is the use of weak lenses and narrow beams. Apertured lenses appear to have more spherical aberration than any other type of focusing lenses. This effect results from the dipping of the field into the apertures which results in nonuniformity of the lens. The cylindrical lens has proved to be the best electrostatic lens for minimum aberration.

PERFORMANCE CHARACTERISTICS

ELECTRON-GUN REQUIREMENTS AND FOCUS CHARACTERISTICS

The design of an acceptable electron gun is fundamentally an optimum compromise of a number of interrelated parameters. As one characteristic is enhanced some other favorable characteristic will probably have to be sacrificed. It follows that a great deal of experience is required for gun design, and that the most critical designs tend to be evolved after a series of design and field tests.

Aside from the effects of lens errors, the characteristics of a lens that are of most interest from an applications standpoint are the depth of focus and sensitivity of the focus to variations in grid No. 2 voltage, ultor voltage, and beam current. The depth of focus is a measure of the ability of the lens to remain in sharp focus as the focusing voltage is varied from the optimum value. Depth of focus may be defined as the difference between the minimum voltage at which a just perceptible loss of picture quality can be seen and the maximum voltage that will produce a just perceptible loss of quality. (The "just perceptible loss of picture quality" will correspond to about 100 lines loss of resolution.) Depth-of-focus characteristics for typical

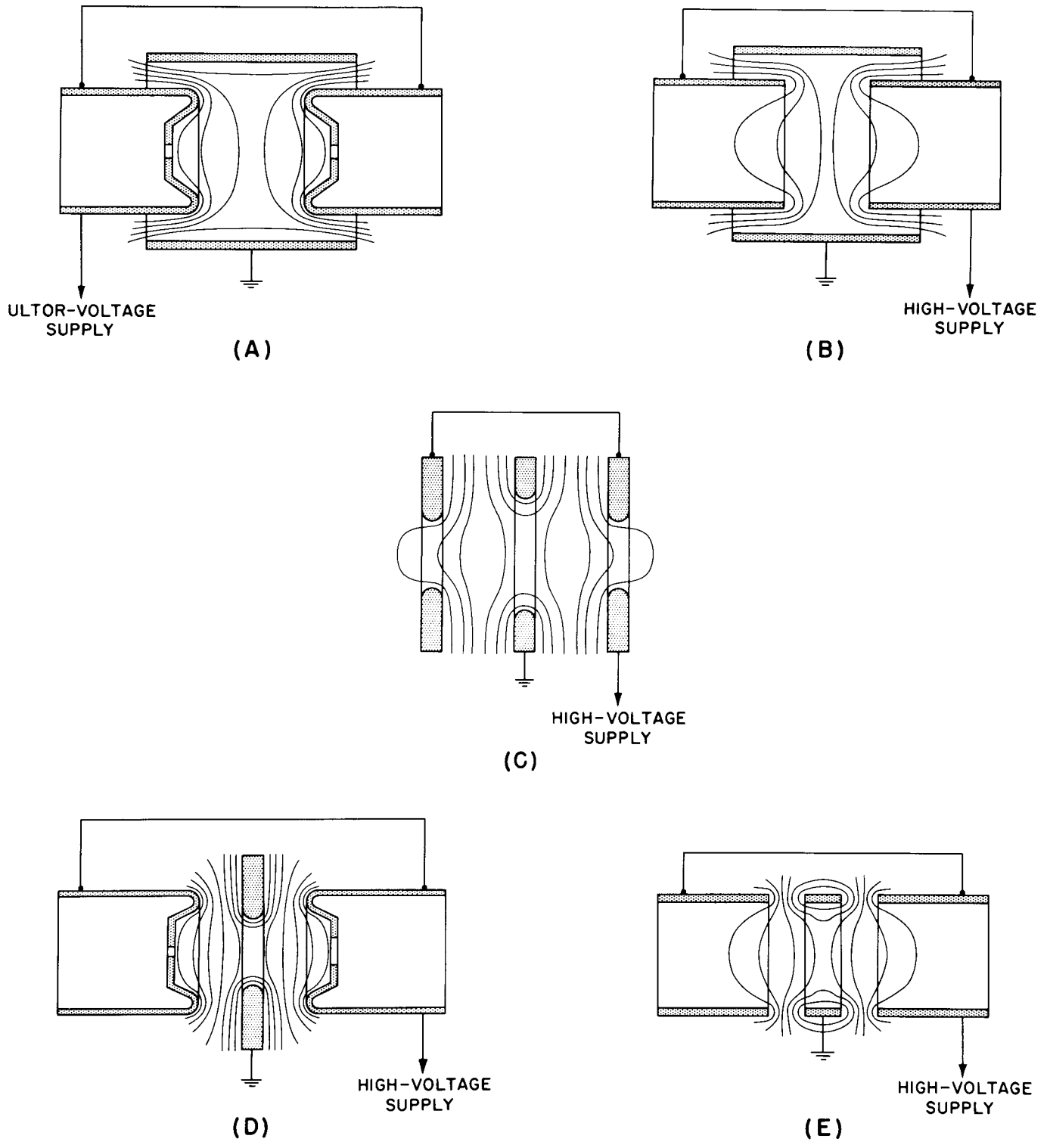


Figure 10. Electrostatic, Einzel-Type Focusing Lenses

picture tubes using an Einzel lens and a low voltage electrostatic acceleration lens are shown in Fig. 11.

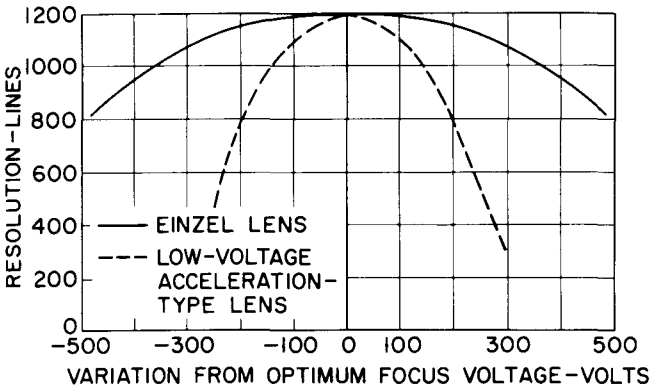


Figure 11. Depth of Focus Characteristics of Einzel Lens and Low-Voltage, Acceleration-Type Lens

Typical depth of focus for an Einzel lens is 400 volts; for a low-voltage electrostatic acceleration lens, it is 200 volts. A depth of focus above 300 volts will insure the consumer sharp focus independent of line voltage and chassis component variations and, as a result, will for the equipment manufacturer eliminate the need for a focus voltage potentiometer and critical focus adjustment.

Focus sensitivities are measures of the changes required in the optimum focus voltage as a function of the operating parameters (grid-No. 2 voltage, ultor voltage, and ultor current). Focus sensitivities of typical kinescopes using Einzel lenses and low-voltage electrostatic acceleration lenses are shown in Figs. 12A, 12B, and 12C.

Modern picture tubes are required to produce distinct and sharply defined scanning lines at both low and high beam current. To meet this requirement, spot diameter should be about 0.040 inch at low drive values, and approximately 0.100 inch at high drive values. In addition, the ratio of spot diameter at the center of the screen to spot diameter at the edge should be between 0.6 and 1.0.

PICTURE QUALITY

Picture quality is the measure of the ability of a picture tube to faithfully reproduce an optical image or picture from an electrical input signal. It may be evaluated both subjectively and objectively. Subjectively, picture quality may be evaluated by judging a test pattern or other picture in terms of: (1) tone gradation (gray scale rendition), (2) graininess (presence of "snow" or of screen nonuniformities), and (3) sharpness (snap or resolution). Objectively, picture quality is determined by measuring such kinescope characteristics as: (1) transfer characteristics (drive curves), (2) spot characteristics (spot size and intensity distribution), and (3) contrast radio. Each subjective measure depends upon a combination of objective measures. [14]

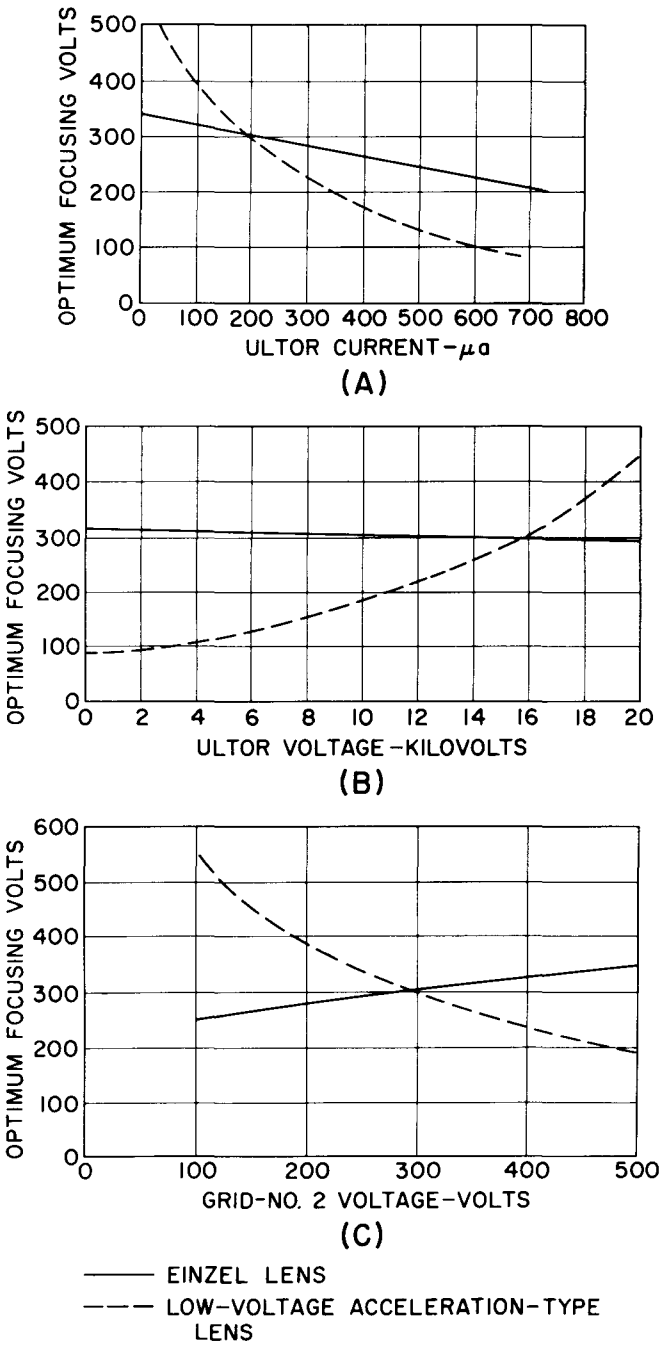


Figure 12. Focus Sensitivities of Einzel Lens and Low-Voltage, Acceleration-Type Lens — variation of optimum focusing voltage with: (A) ultor current; (B) ultor voltage; (C) grid-No. 2 voltage

Subjective Evaluation of Picture Quality

Subjective testing can be used to obtain the customer's point of view. For example, several tubes to be tested can be placed side by side and operated under similar conditions. Observers, through personal judgment and opinion of the test pattern or scene, evaluate the tubes for each of the subjective measures. This method is certainly important; however, it suffers inaccuracy due to human limitations, opinions and judgment. Objective

measures are more absolute, but are sometimes difficult to interpret in subjective terms.

Factors Determining Picture Quality

Good picture quality is obtained by optimizing the gun design parameters so that the electron beam has the desired shape. Among other things, the distribution of electrons in the crossover region and the action of the main focusing lens on the beam determine: (1) the size, (2) the shape, and (3) the intensity distribution of the optical spot on the screen. These spot characteristics, in turn, determine the picture sharpness. Since the spot cannot possibly reproduce lines that are narrower than the spot diameter, the basic size or diameter of the spot limits the resolution of the picture. The designer strives for minimum spot diameter compatible with compromises involving light output, reliability, and manufacturing cost. Spot diameter increases directly with beam current. Under extreme high-drive conditions (near zero bias voltage) the aberrations of the crossover forming lens may be sufficient to form a relatively large-diameter halo, or bright disk, superimposed upon the spot. This halo may contain a significant percentage of the total light flux or energy. In addition, light reflection from the front glass surface of the tube face produces a circular disk of light around the spot on the screen surface. This disk is called optical halation. The size of this halation disk is determined primarily by the thickness of the faceplate. The relative brightness of the halation disk is determined by the color of the screen material, the transmission of the faceplate glass and the surface condition of the front of the faceplate.

Ellipticity of the spot may cause resolution in the vertical directions to differ from that in the horizontal direction as a result of the difference in effective spot diameter. Ellipticity can be caused by misalignment of the gun electrodes, asymmetrical fields near the beam (e.g., "bead charging") and by deflection defocusing, which is apparent at the edge and corners of the screen.

The distribution of intensity across the spot has the greatest effect on picture sharpness. As in light optics, the distribution of brightness or light intensity across the diameter of the spot approximates a cosine-squared curve. If r is the radius of the spot and I is the intensity at the center of the spot, the intensity at a distance d from the center is $I\cos^2(d/r)$.

In practice, the intensity distribution of a spot deviates from the cosine-squared function. Intensity distribution determines the sine-wave response of the tube and can be determined by measuring sine-wave response.

Objective Measurement of Picture Quality

There are three methods of objectively measuring spot characteristics (size and intensity distribution). These methods are: (1) measurement of spot diameter, (2) measurement of limiting resolution, and (3) measurement of sine-wave response. The first two methods are widely used; the third, because it requires the use of highly specialized equipment, is seldom used. It is, however, becoming more important and valuable as engineers learn to apply and interpret response data.

Measurement of Spot Size. The simplest method of evaluation is measurement of spot size using a dot generator and a microscope with a calibrated reticle. The output pulse of the generator is adjusted to drive the tube to predetermined peak current levels in the dots, while the remainder of the screen is kept below cutoff to eliminate background light.

Measurement of Limiting Resolution. The second method of evaluation, although somewhat subjective and inaccurate, is the measurement of the "limiting resolution" of a kinescope. This is done by applying an "Indian Head" pattern, or other calibrated test pattern to the tube. By locating the point on the pattern where the lines in the vertical or horizontal wedges are no longer distinguishable, a measure of the limiting resolution is obtained.

Measurement of Sine-Wave Response. The third and most accurate method for determining spot characteristics is the measurement of the sine-wave response of the kinescope.[14, 15] The first step in this method is to scan the tube in a manner such that a vertically compressed raster is produced. The raster is compressed to the point where the scanning lines overlap and are not individually distinguishable. An audio-frequency sine-wave signal is then imposed upon the driving electrode at a frequency that will cause the resulting horizontal bars to drift vertically across the screen. The light intensity in a minute area of the screen is then measured by means of a photocell behind an optical system containing a very narrow aperture or slit. As the sine-wave pattern passes across the area being observed, the sinusoidal fluctuation in light level through the slit generates a sinusoidal photocell current. The amplitude of this fluctuation in current is recorded. The frequency of the imposed sine-wave modulation is increased, causing the bars on the screen to decrease in width and spacing. As the width of the bars decreases, eventually approaching that of the picture tube spot, the amplitude of light fluctuation decreases due to the inability of the spot's intensity distribution to accurately reproduce the sine-wave form.

Plotting the amplitude of the current fluctuation against the frequency of the applied signal (converted to lines-per-inch on the screen) gives the sine-wave response curve. Fig. 13 shows the sine-wave response of a typical 23-inch, 114-degree kinescope. The sudden drop in response of about five per cent at about one line per inch is of interest. This drop is caused by the presence of the optical halation disk, and can be used as a rough measure of the size and relative intensity of the disk.

The response curve can be transformed by Fourier analysis into the intensity distribution of the spot. This intensity distribution is of great value to the design engineer in the measurement of aberrations. Because it is difficult to directly interpret the sine wave response in terms of picture sharpness, it is seldom used. The

sine wave response can be used to determine a measure of equivalence, or figure of merit, by a method developed by O. H. Schade, Sr.[15] Schade's figure of merit N_e is the "equivalent optical passband" of constant amplitude containing the same total sine-wave energy as the actual passband (sine-wave response) of the spot being measured.

The equation of N_e is:

$$N_e = \int_0^\infty N(r\psi)^2 \, dN$$

where $r\psi$ is the sine-wave response factor (normalized to 100 per cent at zero lines per inch) at line number N. N is the number of black and white lines generated by the modulating signal. In other words, the area under the squared normalized response curve can be represented as a rectangle (Fig. 14) of constant 100 per cent amplitude response from N = 0 to N = N_e. At N_e the response of the equivalent optical passband drops abruptly from 100 per cent to zero.

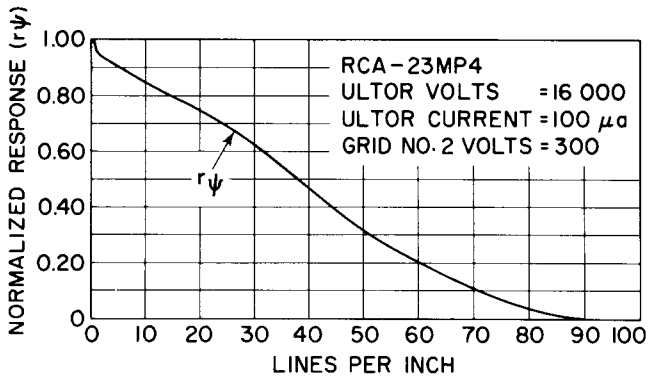


Figure 13. Sine Wave Response of RCA-23MP4

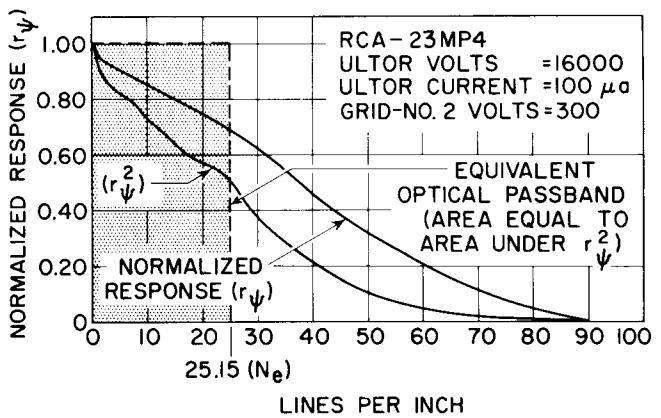


Figure 14. Equivalent Optical Passband of RCA-23MP4

Photographic tests made with a number of pictorial subjects, including test patterns, have shown that the relative definition of images with equal N_e ratings is substantially equivalent even though the limiting resolution may differ by a factor of as much as three.

Furthermore, the sharpness of an over-all television system (including lenses, photographic films, amplifiers, transmission lines, etc.) may be determined by measuring the N_e of each component part and calculating the total from the following equation:

$$\frac{1}{N_{e(total)}} = \left[\left(\frac{1}{N_{e_1}}\right)^2 + \left(\frac{1}{N_{e_2}}\right)^2 + \cdots + \left(\frac{1}{N_{e_n}}\right)^2\right]^{1/2}$$

Since the response of the human eye is known in terms of N_e at particular viewing distances, the above equation, by the addition of an N_e term for the eye, can also account for the eye's response and be used to determine the entire system response.

By correlating the objective measure N_e with a subjective measure of sharpness, the effectiveness of a change in any of the components of a television system can be evaluated.

A subjective unit of sharpness, the liminal sharpness unit, has been defined by Baldwin.[16] He states that one liminal sharpness unit (limen) is a barely perceptible sharpness difference. For use in comparative tests, the liminal unit is designated as the difference in sharpness which will result in 75 per cent of a group of viewers judging one picture to be sharper than another. Greater differences in sharpness can be similarly related statistically to viewer preference. The basis of Baldwin's scale in shown in Table I.

Table I

Difference in Limens	Per Cent of Observers Judging One Picture To Be Sharper	Remarks
0	50	50% guess wrong
1	75	Just perceptible (50% know 50% guess: 25% right 25% wrong)
2	91	Noticeable
3	98	Somewhat Objectionable
4	99.5	Objectionable

Fig. 15 correlates $N_{e(total)}$ multiplied by the screen height V with the subjective measure of liminal sharpness. Fig. 15 and Table I can be used to predict the viewer's subjective judgment of sharpness from the measured sine-wave response of the picture tube. As a result, the sine-wave response measurement is extremely useful in the evaluation of relative performance and picture quality.

RELIABILITY

Reliability is the probability that a tube will perform satisfactorily for some fixed period of time.

Reliability is extremely important in the picture tube industry because of the practice of giving a comparatively long-term warranty and of the major economic considerations involved. In addition, long tube life is

essential to insure customer satisfaction, and thereby, to maintain the tube manufacturers' competitive position in the market.

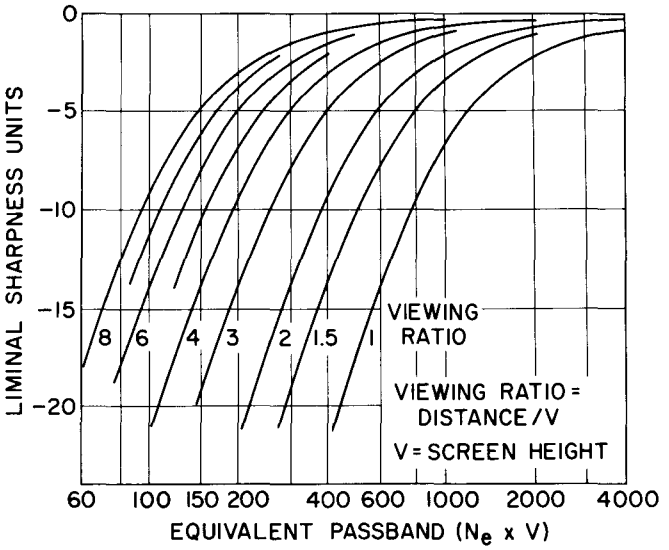


Figure 15. Liminal Sharpness Units as a Function of Equivalent Passband for Several Values of Viewing Ratio

Picture tube reliability may be affected by the design of the product, by the control of the basic reliability parameters during tube manufacture, and by the conditions under which the product is used. In general, major reliability improvements are realized when the designer advances the state of the art or develops more sophisticated techniques. For example, the use of an improved large-area getter has doubled the average emission life of picture tubes; the development of improved materials such as a tungsten-nickel cathode has also been used to double the average emission life of kinescopes.

For a given design, the important reliability factors must be adequately controlled to minimize the economic risks involved in short-term epidemic problems. Sound statistical control imposes a formidable economic problem due to the large sample size required to obtain the required low rejection level at a high confidence level. For example, a sample of 750 is required to detect a one-half per cent rejection level with 95 per cent confidence. Practically, techniques are developed to control the material, process, and manufacturing variations separately and to evaluate all internal and external data sources for deviation trends in order to obtain adequate economic control.

From the standpoint of cost, the optimum reliability of the product is that which will minimize the sum of the manufacturing cost and the cost of in-warranty replacement. The engineering objective is to optimize the reliability cost. Ordinarily, increasing the reliability of a product will increase its cost. However, it is of interest to note that the major picture tube reliability improvements have been achieved by increasing the sophistication of the design without increasing the cost significantly.

The picture tube failure-rate characteristic approximates a skewed cumulative gaussian distribution and is determined by a large number of basic independent and interrelated parameters. Failures may be divided into two types: catastrophic and wear-out. Catastrophic failures are sudden, complete failures such as those that result from air leaks, open heaters, or shorted electrodes, and generally occur early in life. Catastrophic failures are controlled and minimized by material testing, process controls, and short-term life testing at maximum rated operating conditions. Wear-out failures result from gradual deterioration of characteristics, such as might occur as a result of loss of emission, to the point that the tube will no longer meet standards for acceptable performance.

The criteria for acceptable performance of picture tubes appear to depend upon warranty period. Consumers tend to reject tubes for relatively minor defects during the warranty period, whereas, tubes are rejected only for major deterioration of performance or catastrophic failure after the warranty period. Wear-out failures tend to occur later in life and are controlled separately by material testing, process controls, analysis of failures, and long-term life testing under typical operating conditions.

The average emission life expectancy of picture tubes ranges from three to eight years usage depending upon the manufacturer. This long life is required in order to minimize the cost of failures during the warranty period and to maintain competitive customer acceptance.

Picture tubes have in-warranty failure rates ranging from three to five per cent depending upon the manufacturer. This rate includes catastrophic failures and wear-out failures, as well as the tubes rejected during installation in the receiver by the equipment manufacturer for borderline performance characteristics.

TRANSFER CHARACTERISTICS

The drive, or modulation, characteristics of picture tubes are normally presented in the form of a transfer characteristic. Fig. 16 shows a transfer characteristic of a typical picture tube.

For grid drive, the control grid is normally biased to cutoff with a negative voltage, the cathode is grounded, and a positive video signal voltage is applied to the grid. For cathode drive, the grid is normally operated at or near ground potential, the cathode is biased to cutoff with a positive voltage, and a negative video signal voltage is applied to the cathode.

It is apparent from Fig. 16 that the modulation sensitivity of a kinescope is higher when it is operated under cathode-drive service than it is when operated under grid-drive service. This is due to the difference between the dynamic relationship and the static relationship of the potential difference between grid No. 2 and the cathode. Most television applications use cathode drive due to circuit economies.

The modulation sensitivity can be enhanced for kinescopes used in cathode-drive service (as shown in

Fig. 17) by designing the gun for operation at low grid-No. 2 voltage. In such a design, the spacing between grid No. 1 and grid No. 2 is reduced to about 0.005 inch, the grid-No. 1 aperture is increased from 0.022-0.025 inch to 0.031-0.036 inch, and the spacing between grid No. 1 and the cathode is decreased to provide about a 45-volt cutoff voltage with a grid-No. 2 voltage of 50 volts.

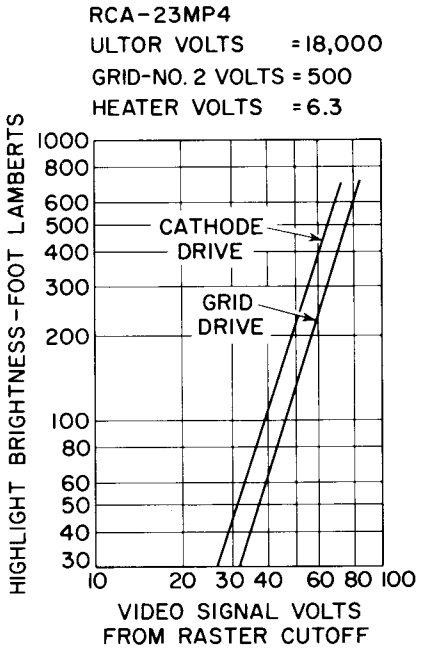


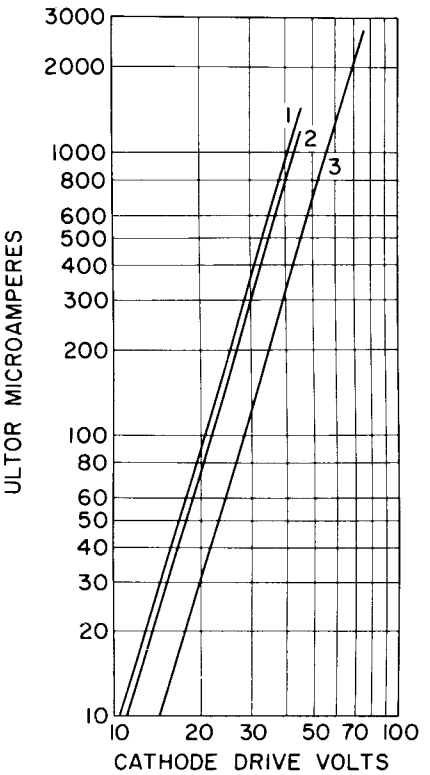
Figure 16. Transfer Characteristics of RCA-23MP4

A kinescope for operation at low grid-No. 2 voltage provides the advantage of reduced drive voltage requirements and related circuit economy, reduced background picture agitation, and improved apparent contrast for a given drive signal. These advantages can be achieved only by accepting the following inherent disadvantages: increased spot size with the resulting loss of picture quality, increased spot blooming, and limited peak highlight brightness unless the tube is operated out of rating for peak negative cathode to grid-No. 1 voltage. The resulting compromise has greatly restricted the use of this type of picture tube design.

The brightness of a picture tube is determined by the ultor voltage and current. The design of each type of picture tube is optimized for the intended operating conditions. Fig. 18 shows the general relationship between the ultor voltage and beam current of a typical kinescope.

GUN EFFICIENCY

Gun efficiency and the resultant highlight brightness are a function of beam diameter and beam centering since the small apertures in the portion of the gun nearer the screen have to be cleared; otherwise, serious beam masking will take place and the curve for beam current (Fig. 18) will show saturation at high values. In determining the efficiency requirements of a gun design, it is first necessary to ascertain if performance objectives require masking of the outer extremeties of the beam. The required diameter of the masking aperture can be determined when the beam diameter, masking aperture position, and the amount of masking required are known.



CURVE 1: 23NP4 WITH GUN DESIGNED FOR OPERATION AT LOW GRID-NO. 2 VOLTAGE
ULTOR KILOVOLTS = 16
GRID-NO. 2 VOLTS = 50
CUTOFF VOLTS = 46
HEATER VOLTS = 6.3

CURVE 2: 23MP4 CONVENTIONAL GUN
ULTOR KILOVOLTS = 16
GRID-NO. 2 VOLTS = 230
CUTOFF VOLTS = 46
HEATER VOLTS = 6.3

CURVE 3: 23MP4 CONVENTIONAL GUN
ULTOR KILOVOLTS = 16
GRID-NO. 2 VOLTS = 445
CUTOFF VOLTS = 75
HEATER VOLTS = 6.3

Figure 17. Modulation Sensitivity of a Picture Tube with Low Grid-No. 2 Voltage Compared with that of a Conventional Picture Tube

Another important factor that affects the efficiency of picture tube guns is the accuracy of the alignment of the beam with the apertures, particularly beyond the triode section of the gun where the beam diameter is large. The accuracy is determined largely by the basic gun alignment, that is, the alignment of the beam-forming and accelerating apertures in the portion of the gun

nearer the cathode. Misalignment of these apertures results in a tilted beam with respect to the geometric axis and the apertures of the portions of the mount nearer the screen. In general, a 0.001-inch aperture misalignment tilts the beam off axis by one-fourth degree.10 The result is partial beam masking by the apertures due to the miscentered beam. Any masking represents a direct reduction in brightness since the light output of a kinescope is directly proportional to the current exciting the phosphor. It is advantageous with certain designs to mask the beam to reduce the spot size, reduce the effective beam diameter, increase the resolution, or improve the spot shape, especially for high beam current values.

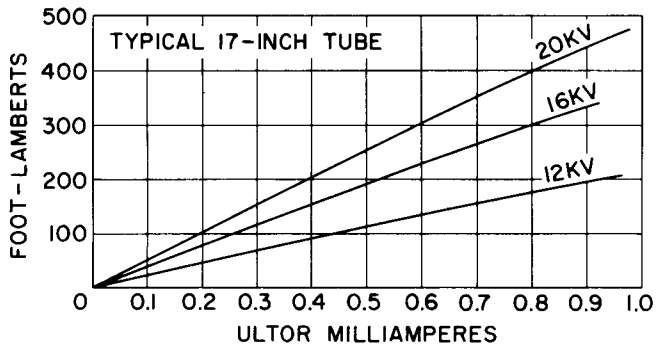


Figure 18. Brightness of a Typical Picture Tube as a Function of Ultor Current for Several Values of Ultor Voltage

TUBE DESIGN CONSIDERATIONS

BULB DESIGN

The bulb of a cathode ray tube serves as the vacuum-tight member enclosing the electron gun and the phosphor screen. This bulb may be made completely of glass or may be a metal-and-glass envelope with two glass-to-metal seals. Currently, all-glass types are made in the largest volume, although following World War II, the metal-glass bulb enjoyed periods of very high production.

A typical metal-glass bulb consists of four parts: (1) a uniformly thin curved faceplate of glass, (2) a metal cone, (3) a glass funnel, and (4) a flared neck section as shown in Fig. 19.

The faceplate is produced in a sheet, cut to size, and then sagged to give it the proper curvature. The metal shell is spun from a flat sheet and finished with sealing lands at each end. The funnel is pressed with a thin glass wall in the region under the deflection yoke and is sealed to the flared neck section. The funnel-neck assembly is first sealed to the metal cone and then the faceplate is sealed to the assembly. Two materials have been used with success for the metal cone. These two are an alloy of chromium and iron (either 37 per cent or 17 per cent chromium) and cold-rolled steel. Different glasses with proper matching thermal expansion coefficients are required for these metal cones.

A typical all-glass bulb (Fig. 20) also consists of four parts: (1) a pressed face panel, (2) a pressed or spun funnel, (3) a contact button, and (4) a flared neck section. The all-glass bulb is assembled by first sealing the button in the funnel and then sealing on the flared neck by means of gas-air burners. Next, the face panel is sealed to this assembly by an electric-arc technique, where the thick glass walls are heated uniformly throughout by high-frequency electrical power to make a strong seal.

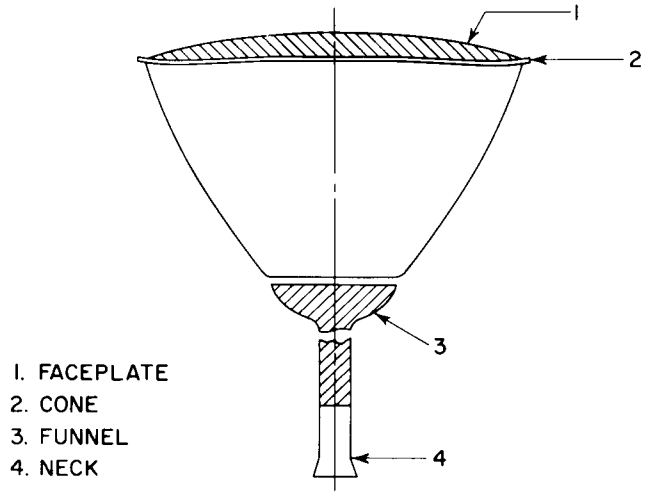


Figure 19. Glass-Metal Bulb Assembly

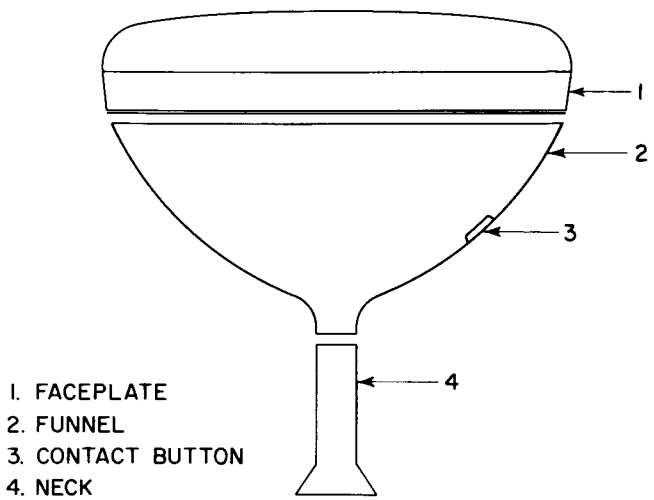


Figure 20. All-Glass Bulb Assembly

The glass-metal type of bulb served a very useful role in the development of modern black-and-white picture tubes. Because of it, vendors producing the all-glass bulbs were required to develop improved processes to produce high optical quality in the pressed face panel, lightweight face-panel designs, and lightweight funnels, having thin wall thickness in the region under the deflection yoke. With the development of the improved glass processes and high-volume production equipment, the all-glass bulb was able to overcome the inherent quality and economic advantages of the metal-glass type and eventually to replace it.

The metal-glass bulb still enjoys a weight advantage

over comparable size all-glass bulbs because of its thin uniform faceplate and its lighter weight, stronger metal shell. It presents somewhat less of a safety hazard when opened to air accidentally because an implosion will not necessarily result. An implosion, however, is somewhat less hazardous because fewer glass particles are produced and the metal cone tends to confine them. Since the glass for the faceplate is drawn in sheets, the plates can be cut so as to avoid defects in the sheet, thus giving them high optical quality. The metal-glass bulb also has the advantage of thinner glass in the yoke region and, hence, less interference between the deflected beam and the inside glass surface. The metal-glass bulb has the disadvantages of requiring a larger cabinet to provide adequate electrical insulation between the metal shell, which operates at ultor potential, and the metal parts of the cabinet and chassis; and of possible separation of the glass-metal seals due to temperature changes and shock.

The size and shape of the bulb for cathode ray tubes has undergone a continuous evolution during the period following World War II. Table II lists the tube types typical of production during this period and tabulates pertinent defining data.

These data show the evolution from a 10 inch tube with a deflection angle of approximately 50 degrees to the modern 23 inch 114 degree tube with a net decrease in over-all length. During the intervening period, the over-all length has remained about constant. As the demand for larger picture sizes has developed, these types have been designed with wider and wider deflection angles and shorter neck lengths. The maximum over-all length has been influenced by styling considerations and the influence of doorway sizes on television set dimensions. The change to rectangular tubes was prompted by consumer preference, the desire to save cabinet space on the larger-size tubes, and the desire for enhanced styling. Inherently, the rectangular bulb is weaker because of its shape. The ideal bulb shape for strength with minimum weight is a sphere. Because of the difficulty of achieving a true rectangular shape without excessive glass weight, the evolution of more nearly rectangular bulbs has been made in steps. The tubes listed as rectangular are actually rectangles with rounded sides and corners. The "new look" tubes closely approximate a rectangular shape.

Faceplate curvature affects bulb strength. Ideally, the face should be curved as much as possible and

Table II

Tube Type	Design	Year	Envelope round or rectangular	Screen Width inches	Picture Area square inches	Maximum Over-All Length inches	Diagonal Deflection Angle degrees
10BP4	Glass	1946	Round	9-3/8	-	18	52
12LP4	Glass	1949	Round	11-3/8	-	19-1/8	57
16AP4	Metal	1949	Round	14-3/8	-	22-5/16	53
16GP4B	Metal	1950	Round	12-3/8	-	17-11/16	70
17CP4	Metal	1950	Rectangular	14-5/8	146	19	70
21AP4	Metal	1951	Rectangular	18-3/8	230	22-5/16	70
21AMP4A	Glass	1954	Rectangular	19-1/16	262	20-3/8	90
24CP4	Glass	1954	Rectangular	21-7/16	332	21-1/2	90
21CEP4	Glass	1956	Rectangular	19-1/16	262	14-3/4	110
24AHP4	Glass	1956	Rectangular	21-7/16	332	16-3/16	110
23CP4	Glass	1959	Rectangular*	19-5/16	282	15-9/16	110
19AFP4	Glass	1960	Rectangular*	15-1/4	172	11-15/16	114
19XP4	Glass	1960	Rectangular**	15-1/8	172	11-5/8	114
23MP4	Glass	1960	Rectangular**	19-1/4	282	14-11/16	114

*"New-look" tube with laminated implosion panel
**"New-look" full-strength tube (more nearly rectangular shape)

should have nearly equal inside and outside radii. This produces a facepanel of uniform thickness which has the advantage of being strong under pressure loading as well as being thermally strong. The "new look" tubes have relatively flat facepanels with a composite curvature to control the thickness as required for thermal and mechanical strength.

The curvature of the funnel of the bulb has its effect on bulb strength. To produce a strong bulb, it should be designed as curved as feasible with no abrupt changes in curvature and a minimum of flat surfaces. Bulb strength may be tested by pressure testing to failure, or by strain-gauge analysis as the bulb is pressure loaded. Typically, bulbs for cathode ray tubes are expected to withstand an absolute pressure differential of 55 pounds per square inch minimum unabraded and of 45 to 50 pounds per square inch minimum after abrasion with No. 150 grit.

Faceplate curvature depends on many factors, of which bulb strength is only one important consideration. The faceplate curvature is influenced by need to provide good uniformity of focus and suitable raster shape. If maximum bulb strength were the sole consideration, that factor would dictate a faceplate of extreme curved contour; if, however, styling and customer preference only were involved, the choice would be a flat faceplate. As a practical matter, focusing and raster shape must also be considered; these factors require a curvature between the two extremes, ideally, with a radius equal to the distance between the inside surface of the faceplate and the apparent center of deflection of the beam. As a general rule, the actual radius chosen is somewhat longer than that required for best focus, but it provides a more nearly flat faceplate with some disadvantages due to nonuniform focus across the raster and pincushioning of the raster shape.

Fig. 21 shows an ideally deflected beam in a tube with a spherical face having a radius of curvature equal to the distance between the deflection plane and the inside surface of the sphere. It also shows the effect of a flat viewing surface that is tangent to this spherical face at the center. The plan view shows the pattern observed when the beam is scanned onto spherical and flat surfaces with equal deflection power applied. With the deflection currents adjusted so that equal screen width and height are obtained on the scanning axes of the spherical surface, the pattern reproduced is square; whereas, the picture on the flat surface is distorted as shown. The elevation view shows that, since the electrons are moving in a straight line, the longer the path length for a given angle θ the greater the radial displacement. It is also quite apparent that as the angle of deflection θ increases, the change in the displacement increases and gives a pincushion pattern with the increased area as is shown in the plan view for the flat surface. Similarly, since the gun has a fixed focal length (for a given focus setting), the beam will become increasingly defocused as it is scanned from the geometric center onto a flat plate. In practice, pincushion and focusing defects can be minimized by changes in the winding distribution of the deflecting coils. Raster shape can also be corrected by permanent magnets. However, improvement of one defect by such means usually results in degradation of the other. It is, therefore, imperative that the faceplate radius approximate the ideal value, as closely as possible. Bear in mind that as the faceplate curvature is increased for a given screen dimension and deflection angle, the bulb length is also increased.

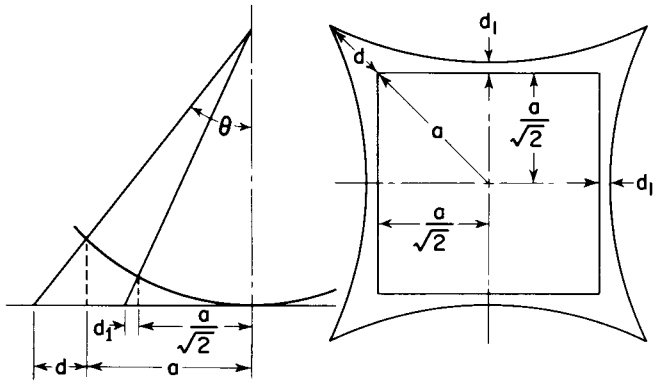


Figure 21. Comparison of Flat and Ideal Spherical Picture-Tube Faces

Modern rectangular tube designs are called "new look" types (see Table II). The 23 inch 110 degree bulb has essentially the same glass thickness as used in the conventional-design tube and as a result has reduced strength. The facepanel of this 23 inch tube is protected by a safety cap laminated to its outer surface by an epoxy resin. This cap covers the high stress area of the bulb and prevents weathering of and surface damage to the weak area of the facepanel. In addition, the resin tends to hold the facepanel together when the bulb is broken, prevents flying glass, and thus results in nonviolent implosions. The 23 inch 114 degree bulb has been designed to have sufficient strength without the laminating cap. Thickened glass sections are used in the high-stress areas of the sidewalls of the facepanel. This type of tube can be mounted conventionally and used with a separate safety glass instead of the laminated safety cap.

The 19 and 23 inch tube designs feature improved large-area contrast as compared with rectangular tubes of conventional design. The improvement results from the additional scanned picture area provided by the more-nearly rectangular corners. For prototype designs, more of the picture corner information is scanned into the funnel sidewalls. The resulting electron diffusion causes loss of contrast by producing general screen background illumination.

Conventional and laminated bulbs utilize a filter-glass for the panel. This glass is made from a modified glass composition that contains a neutral gray filter additive. The function of this filter is to reduce the contrast dilution effect of ambient lighting and to reduce optical halation. This effect is achieved because about 25 per cent of the light is absorbed each time the light passes through the glass. It is recognized that the ambient and halation light passes through the filter twice, whereas the viewed light emitted from the screen passes through the filter only once.

The cap, which is laminated to the bulb to produce the bipanel tube, is made from a special filter glass having a 47 per cent transmission. This acts as a series filter and further reduces the halation and ambient light. It is noted that the halo diameter is increased about three times due to the thicker effective panel glass. Conventional tubes are mounted in receivers with a filterglass implosion plate in front of the screen. The conventional tube-implosion plate combination has the same net system transmission — about 40 per cent — as a bipanel laminated tube. The ambient light effect in contrast is practically the same for both types of tubes. The optical halation is reduced somewhat by the thicker constant-index-of-refraction medium and results in a detailed contrast advantage for laminated tubes.[7]

Another important bulb design characteristic is adequate "beam clearance," "penetration," or glass thickness, in the region of the deflecting yoke. The lack of good beam clearance results in the beam striking the glass surface before being deflected enough to reach the edge of the screen. This problem becomes more critical as the deflection angle increases. More and more effort has been directed toward achieving an optimum glass shape and obtaining control of this shape as the angle became successively wider. For the 90 degree types, a set of gauges — JEDEC Nos. 113 and 116 — were designed to measure and control the glass shape. For 110 degree types, it was experimentally determined that the optimum inner contour of the bulb in the vicinity of the yoke should be parabolic. A mica card coated with phosphor and located in the yoke region within the tube was used to obtain a plot of the beam path during beam deflection. The bulb was then designed so that the inner bulb contour matched the beam path in the critical region. The outside shape under the yoke is also parabolic and separated from the inside surface by the minimum glass thickness consistent with process capabilities of the bulb producer and with adequate strength. A set of matching gauges (JEDEC Nos.125 and 126) were also designed to measure the beam clearance of this bulb design.

SCREEN APPLICATION

In the application of fluorescent screens to picture tubes, it is essential that the thickness of the screen be uniform across the face. Uniformity is necessary because the phosphor normally used is a two-component type composed of a white powder emitting blue light, and a yellow powder emitting yellow light. The light being emitted from the tube originates within the phosphor crystals, and is emitted from the tube after multiple scatterings within the screen structure. Since a yellow powder, by the nature of pigment action, does not reflect or transmit blue light, the thicker a screen of given composition, the more yellow the emitted light becomes. Thus, a screen of nonuniform weight produces light of a nonuniform color.

To exhaust tubes properly, it is necessary that binders used for making the screen have low vapor pressure, and binders are thus limited, in general, to inorganic substances. Also, since incorporation of heavy metals into the phosphor may cause changes in emission characteristics, the screen application method must minimize accidental introduction of these elements. To get maximum light output from a screen, it is necessary to have a phosphor structure with a minimum of optical contact between the phosphor and the tube face. Finally, the screen application method should produce complete coverage of the faceplate with the minimum amount of phosphor. A more complete treatment of the optics of the screen structure may be found in Refs. 17 and 18.

The water settling process of screen application best meets the preceding requirements and is universally used. It is described by Hardy in his article in this book on "Application and Colorimetry of Phosphor Screens." The process, although basically simple, must be carefully controlled to assure good screen quality. Heavy-metal contamination of the phosphor is prevented by using demineralized water and high-purity chemicals. Copper and nickel, in particular, are kept at levels of less than 0.01 parts per million in the settling solution. Proper screen weight uniformity is maintained by minimizing eddy currents due to the filling operation; thermal convection currents are controlled by controlling room temperature, glass temperature, and settling solution temperature. The natural tendency of the screen to be heavy at the center of the face due to bulb geometry is overcome by holding solution temperature to somewhat below ambient temperature; in this way, heat transferred to the solution during settling produces convection currents which tend to carry the phosphor from the center toward the edge of the face.

Observation of settling rate of phosphor in bulbs, microscopic determination of phosphor particle size, and consideration of Stokes law, reveal that the phosphor settles in aggregates of individual particles. In order that a relative measurement of the size of these aggregates can be obtained, screens of known weight per unit area are settled on glass, and then measured for light transmission. It can be shown that the coarser the structure, the greater the light transmission for a screen of given weight. The reciprocal of the light transmission, relative to a standard, is used as a measure of effective particle size. A second important settling characteristic is wet adherence, i. e., the adhesion of the phosphor to the glass before decantation of the supernatant liquid. This characteristic is measured by settling a screen under desired conditions, and before decantation, eroding a hole in the screen by impingement of a controlled nonturbulent stream of water. The reciprocal of the eroded hole size, relative to a standard, is used as a measure of wet adherence. Dry adherence, the measure of adhesion after decantation and removal of free water, is measured by impinging a controlled stream of water on a dried screen which has been rewetted.

In adapting the screen-settling procedure to a specific application, consideration must be given to obtaining: (1) optimum light output by choice of the proper effective phosphor particle size and screen weight, and (2) optimum uniformity and adherence of the screen by use of correct quantities and concentrations of water, binder, and electrolyte, as well as of suitable techniques for introducing these into the bulb.

Fig. 22 shows the relationship of phosphor particle size and screen weight to tube ultor voltage and light output. The curves show that for a given size of phosphor particle there is an optimum screen weight, and that this optimum weight increases slightly with increasing ultor voltage. Furthermore, the smaller the size, the lower the optimum weight. It is impractical, however, to take full advantage of the latter observation, because of the long settling times required with smaller phosphor-particle sizes.

Fig. 23 shows the effect of settling-solution concentrations on wet adherence and dry adherence of screens. It can be seen that with respect to wet adhesion, for a given concentration of binder there is an optimum concentration of electrolyte, and vice versa. The same applies to dry adherence.

Fig. 24 shows the effect on screen color uniformity of varying the volume of suspension containing a constant weight of phosphor when the suspension is dispersed onto a settling cushion of volume adjusted so that total liquid volume in the bulb after dispensing is constant.

In practice, a settling procedure for a specific application includes: (1) choosing a phosphor particle size and screen weight to optimize light output for the ultor voltage requirement and available settling time; (2) determining the total cushion volume needed for best uniformity; (3) selecting binder and electrolyte volumes for adequate adhesion; (4) selecting proper phosphor suspension dispensing nozzle for best uniformity, (5) determining the proper water volume to accompany the phosphor-silicate mixture, and (6) (if perfect screen weight uniformity cannot be obtained) matching particle size of blue and yellow components, so that the tendency of thermal convection current to deposit larger particles at the center and smaller particles at the edge will result in good color uniformity by compensating for weight nonuniformity.

FILMING AND ALUMINIZING

About half of the light emitted from a kinescope screen is directed inward toward the funnel of the tube. If not absorbed within the funnel, it may be reflected back through the screen to reduce large area contrast by illuminating black areas. Because this reflection degrades the picture, it is highly desirable to increase the contrast of the picture by depositing a mirror-like surface on the cathode side of the screen. Before this mirror-like surface can be deposited, the porous structure of the screen must be made flat. This is done by saturating the screen surface with water, and laying down on top of the saturated screen a thin, smooth, lacquer. A general description of this process is given in the article by Hardy on "Application and Colorimetry of Phosphor Screens."

Lacquer formulation is a critical part of the filming operation due to the effect on process scrap and lightout of the finished kinescope. Lacquers for spray filming consist of film formers, plasticizers, and solvents. All components must be selected with a view to the completeness with which they may be removed during subsequent bake and exhaust operations without leaving harmful residues. The various methacrylates have

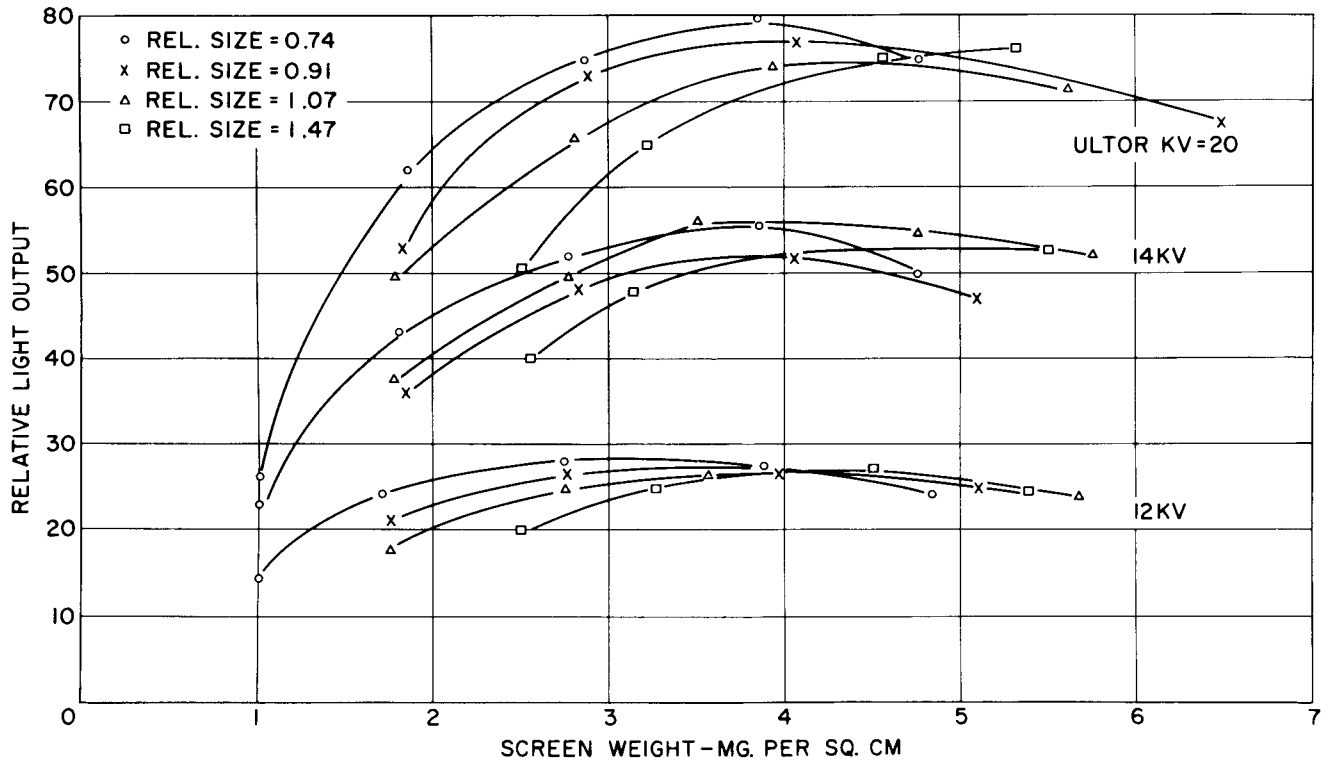


Figure 22. Light Output as a Function of Screen Weight for Several Particle Sizes and Ultor Voltages

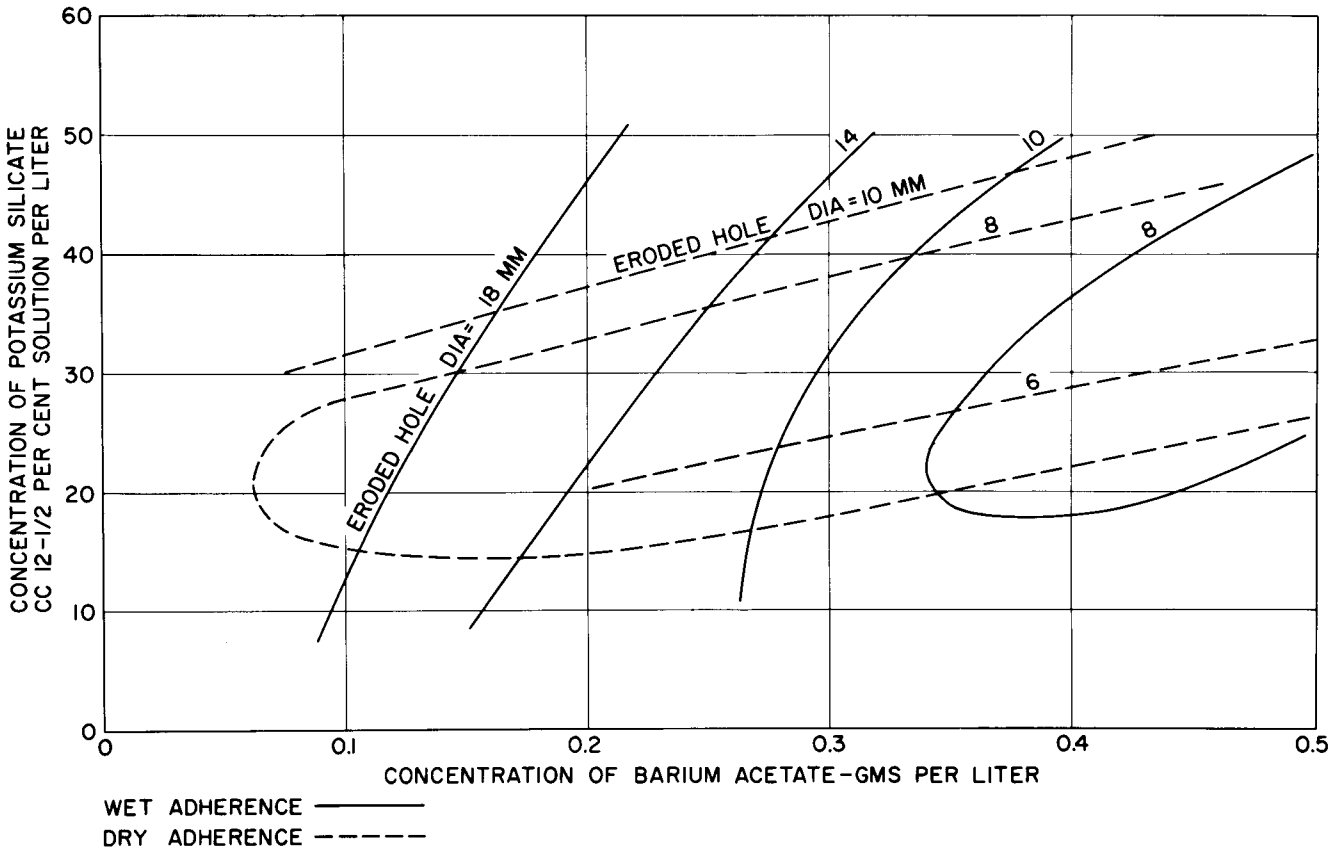


Figure 23. Relative Adherence of Screens

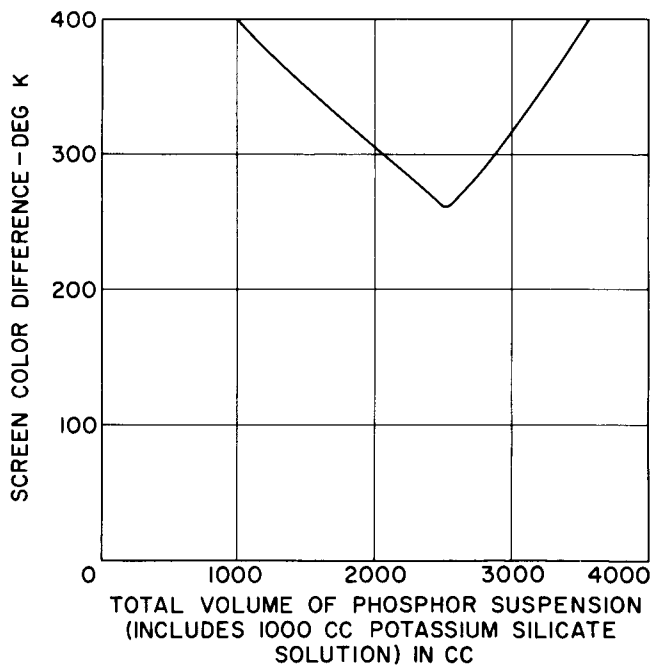


Figure 24. Effect of Phosphor Suspension Volume on Screen Uniformity for RCA-21ALP4

been found to be the most suitable film formers; sometimes cellulose esters are blended with them. Solvents are chosen for proper volatility with evaporation rates in the range from ethyl acetate to toluene. Plasticizers are chosen to give the film the degree of flexibility needed to enable it to cling to the screen without rendering the film so free of gloss that a rough non mirror-like surface results.

The vacuum metalizing process is discussed very completely in Refs. 19, 20, 21, and 22. In general, three conditions must be met. First, the pressure in the vacuum chamber must be such that the mean free path of a molecule at that pressure is approximately as great as the distance from the evaporator to the screen. Second, the material used for the evaporator should be wet by, but if possible, should not alloy with the material to be evaporated. Third, allowance must be made for the fact that aluminum is propagated from the source at a uniform rate per unit solid angle subtended at the source.

In practice, kinescopes are aluminized at a pressure of about 0.1 to 0.5 mm Hg absolute. The evaporation takes place over a period of about one-half to two minutes. A fast pumping system is used because it is necessary to maintain the low pressure condition in the bulb during the evaporation cycle. Careful location of the evaporator is necessary in order to meet the requirements for aluminum thickness. These requirements are: (1) in the deflection yoke region, the aluminum must be thin enough to not shunt-out the yoke field; (2) the aluminum deposited on the screen and funnel must be thick enough to be electrically conductive (about 500 angstroms); and (3) the aluminum thickness

on the screen is, in practice, a compromise between that which gives optimum light output (1000 angstroms) and that needed for ion-spot prevention with straight guns. (Aluminum thickness at the center of the face is usually about 1000 to 3000 angstroms.)

Because of the natural difference in aluminum thickness at the center of the tube compared to the edges of the tube, due to the bulb-evaporator geometry, there is a difference in the electron energy required to penetrate the aluminum at the center as compared to that required at the edges. In addition, because of the angle of incidence of the beam at the edge of the screen, the ultor voltage to light output relationship is not the same at the edge as at the center. This effect is shown in Fig. 25.

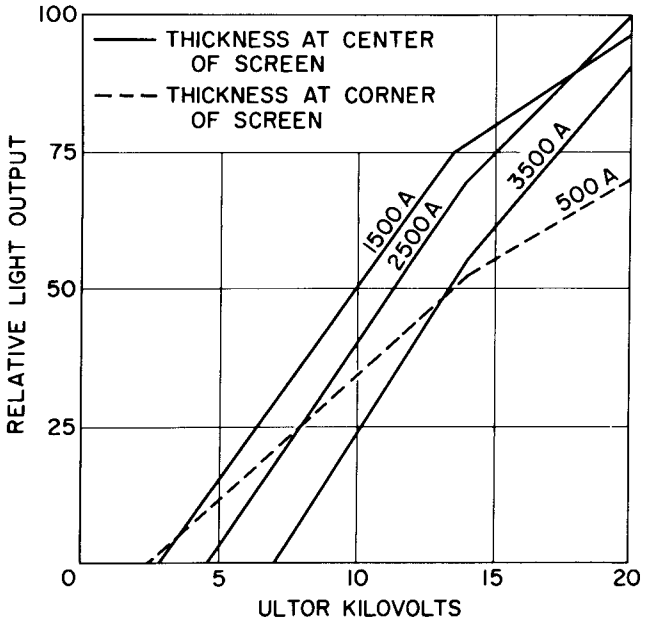


Figure 25. Effect of Aluminum Thickness on Light Output

It can be seen from Fig. 25 that a tube with aluminum thickness of 2000 angstroms at the center and 500 angstroms at the corner, will provide equal light output at the center and corner when the ultor voltage is greater than 6.5 kilovolts. At voltages less than this, the lower light output at the center relative to that at the edge is readily apparent to the eye, and appears objectionable. Thus, as the deflection angle of tubes increases, the minimum anode voltage for proper light output balance increases due to the increasing difference in aluminum thickness between center and edge.

BAKEOUT AND EXHAUST

The principles governing the bakeout and exhaust of kinescopes are similar in most respects to those for other vacuum tubes. The gases enclosed within the envelope and those occluded, absorbed, or loosely combined on the interior surfaces must be removed to such a degree as to assure a favorable cathode environment for adequate life.

The large size of picture tubes places serious limitations on the exhaust process. The rates of heating and cooling are limited by the thermal properties of the heavy-walled soft-glass bulb. With present getters, chemical pumping is not practical and the majority of the available getter must be reserved for use during tube life. The large surface areas of glass, metal, phosphor, and conductive coating that must be completely outgassed, places a limit on the minimum amount of high temperature exhaust bakeout in order to obtain satisfactory tube life. On the other hand, ideal laboratory-type exhaust conditions are impractical for mass-produced picture tubes.

An optimum exhaust process involves many compromises. The parameters governing exhaust must be continually re-evaluated in an attempt to produce high quality, long life tubes economically. Data obtained on the gas content of materials, pressures during exhaust, residual gas content, outgassing characteristics, and optimum bakeout temperatures, combined with life test information, provide the foundation for any kinescope exhaust and bakeout schedule.

Before the exhaust of a kinescope, the coated, screened bulb assembly is lehr baked for a period of about two hours during which time the temperature is kept above 400 C for at least 30 minutes. This preliminary treatment volatilizes the filming lacquer and removes much of the combined and absorbed vapor and gases in the glass, coating, and screen prior to exhaust.

An analysis of the amount and ease of removal of the gases remaining after this treatment has established a number of factors for ideal exhaust conditions: (1) bakeout during exhaust should keep the bulb over 400 C for a minimum of 15 minutes; (2) standing time between lehr bakeout and exhaust should not exceed two hours to prevent appreciable reabsorption; and (3) the quantity of internal conductive coating used should be kept to a minimum to reduce the exhaust removal load.

A comparison of the quantity of gas evolved from typical aluminized and unaluminized tubes, as well as the effect of the addition of coating, screen and mount is given in Fig. 26. An aluminized bulb evolves only about one-fifth of the gas of its unaluminized counterpart. The large difference is attributed almost entirely to the larger quantities of internal graphite coating in unaluminized types.

The effect of holding a bulb 24 hours between lehr bakeout and exhaust on gas content is shown in Fig. 27. The increased holding time results in about a threefold increase in gas content for either type. The gas reabsorbed during extended holding periods is not as completely removed during a 15-minute, 400 C bakeout as that from a bulb held only one hour. By arbitrarily assigning complete gas removal as that quantity of gas removed in 60 minutes at 400 C, a bulb held one hour has only about one-tenth the gas remaining after a 15-minute, 400 C exhaust-bake as one that is held 24 hours (see Table III).

EXHAUST CONSIDERATIONS

An important tool available for the proper selection

of an exhaust index is the pressure curve measured with gauges located inside a tube as it is processed on production equipment. New exhaust schedules proposed for meeting higher production requirements for attaining improved life, or for reducing scrap can be directly compared with proven established schedules. This information, coupled with other data, makes it possible to change processing as well as materials, bulb constructions, and activation schedules with reasonable assurance of the final results.

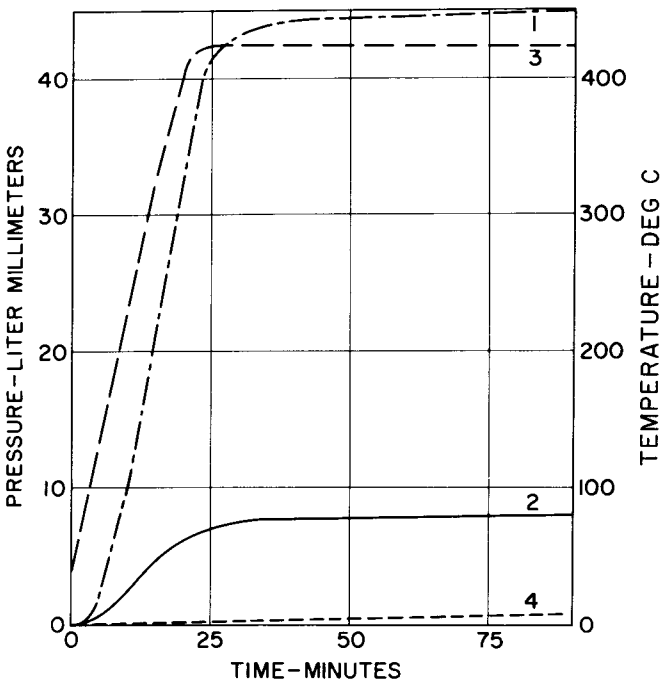


Figure 26. Effect of 1-Hour Holding Time on Gas Content of Picture Tubes: curve 1 — 21ZP4A (unaluminized); curve 2 — 21ZP4B (aluminized); curve 3 — 21-inch, 70° bulb (no screen, mount, or coating); curve 4 — bulb temperature

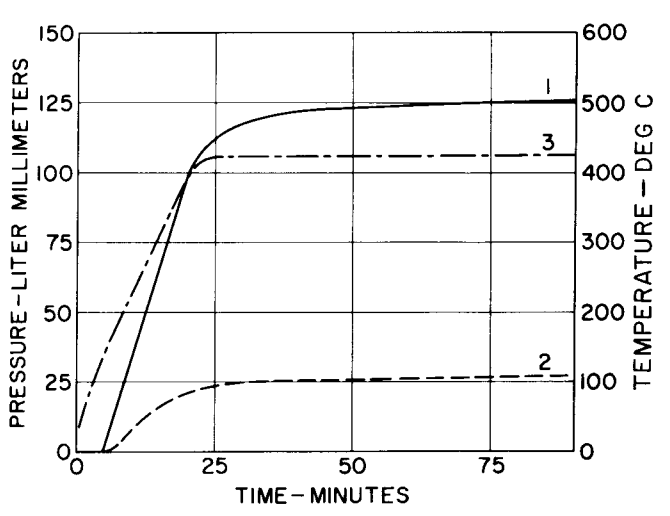


Figure 27. Effect of 24-Hour Holding Time on Gas Content of Picture Tubes: curve 1 — 21ZP4A (unaluminized); curve 2 — 21ZP4B (aluminized); curve 3 — bulb temperature

Table III

Tube Type	Holding Time hours	Gas Remaining liter-mm *			Total Gas† liter-mm
		0 min	15 min	30 min	
21" 70°‡	1	0.5	0.3	0.2	0.7
21ZP4B (alum.)	1	2	0.4	0.2	8
21ZP4B (alum.)	24	12	4	2	34
21ZP4A (unalum.)	1	10	0.5	0.5	45
21ZP4A (unalum.)	24	26	4	2	126

*The time indicated refers to the time the bulb assembly is baked out at temperatures of 400-425 C.

†Total gas arbitrarily refers to that quantity of gas evolved during a 60-minute bakeout at 400-425 C.

‡This tube had no screen, coating, aluminum, or mount. The addition of aluminum coating, screen, and mount to the glass bulb increases the total gas content in an aluminized tube by a factor of ten, whereas, the addition of coating, screen, and mount to the unaluminized counterpart increases the gas content by a factor of 65.

Typical exhaust curves obtained on standard 35-position, straight-line exhaust equipment are shown in Figs. 28 and 29, along with the position of rf outgassing, filament lighting, and grid bombardment. Consolidated MCF 60 diffusion pumps backed by Welch No. 1402 mechanical pumps were used in obtaining these results. The measurements were automatically recorded from ionization gauges using an RCA 1949 gauge tube mounted directly above the grid No. 2 of a modified electron-gun assembly.

The schedule is designed to bring the temperature of grids No. 1 and No. 2 to about 700-800 C without subjecting the hot parts to pressures above 10 microns, and without vaporizing metal onto the glass neck. This temperature, however, cannot be tolerated at the top part of the mount since the glass neck would be broken by the heat conducted through the bulb spacers. On standard 6.3 volt, 600 millampere heaters, peak filament currents of 0.80-0.90 ampere for one to four minutes have proved to give the most satisfactory results. The grid bombardment treatment consists of drawing relatively high current densities (15 to 30 milliamperes) to the grid No. 1 by applying up to +10 volts at the grid. Theoretically, many of the oxides that form the "anode activation spot" on the underside of grid No. 1 are decomposed by 10-volt electrons.[23] The purpose of the grid bombardment process, then, is to break down the oxides and exhaust the gaseous decomposition products from the underside of grid No. 1.

Table IV was compiled from a series of determina-

Table IV

Tube Type	Index seconds	Pressure — microns					Tip-Off Temp—C
		RF	Fil	Grid	Tip-off	Cold	
21" 90° Al	65	2.8	1.2	1.0	.08	.025	280
21" 90° Al	72	1.9	0.7	1.1	.11	.027	280
21" 70° Metal	60	14.0	6.0	2.2	.05	.041	215
17" 90° Al	90	2.8	1.1	.25	.03	.015	185
17" 110° Al	72	6.5	.65	1.7	.06	.021	220
14" 90° Al	40	3.0	.85	.37	.03	—	265

tions on various bulb types. The range of values is considered typical of the product produced during 1956, with known good field reliability. The decrease in pressure between tip-off and the "cold" reading taken 20 hours later is due both to the pressure decrease with lower temperature and to a gettering action by the relatively gas-free internal surfaces. A decrease in temperature from 275 C to 25 C would account for less than a twofold reduction in pressure, so a large amount of the gas remaining at tipoff is actually gettered by the tube surfaces. The effect of decreasing the exhaust tubulation diameter is shown in the results of an experimental 17 inch 110 degree aluminized tube where the inside diameter of the tubulation was 0.175 inch, as compared to the standard size, 0.235 inch. Higher pressures existed at about every point even though the volume of the 110 degree bulb is less than that of the 90 degree bulb. When the design of the 110 degree stem was modified to use a standard exhaust tubulation, results equal to those of the 17 inch 90 degree type were obtained.

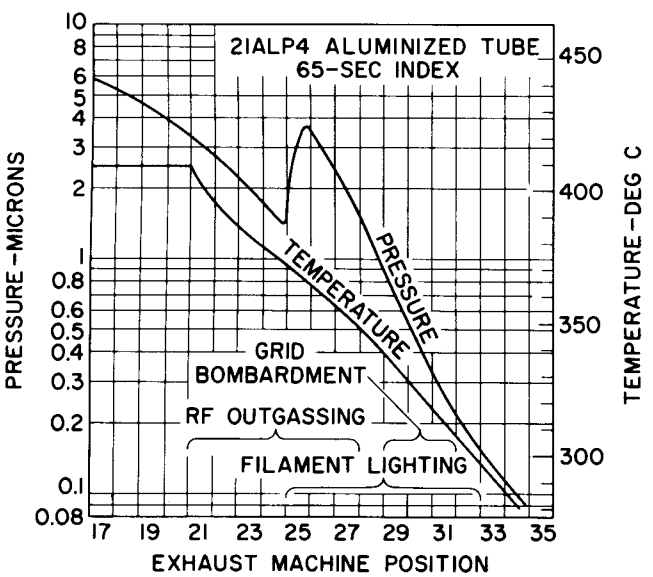


Figure 28. Typical Exhaust Curve for Picture Tube on 35-Position, Straight-Line Equipment

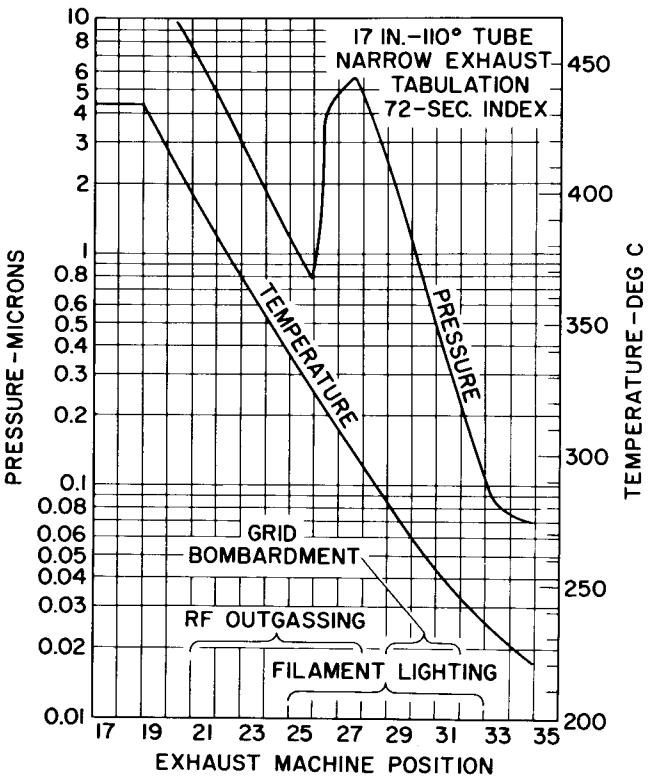


Figure 29. Typical Exhaust Curve for Picture Tube on 35-Position, Straight-Line Equipment

Bulb temperatures at tip-off have been found to be quite critical and should be kept high. Measurements comparing the pressure within the tube to that in the exhaust manifold have consistently shown a cross-over point where the tube pressure falls below the manifold pressure. Since such a condition increases the possibility of pump oil back streaming, the bulb temperature at tip-off is maintained at or above 275 C.

Pressure Stability

The effectiveness of exhaust and of changes in materi-

als can be evaluated by measuring the increase in pressure of an ungettered tube over a period of time. This method has been particularly valuable in comparing various basic bulb-assembly designs.[24] A comparison of three types of metal tube constructions is shown in Fig. 30. Curve 1 is for a metal tube construction with a shell of cold-rolled ND14 steel with both internal and external surfaces of the shell completely covered with enamel; curve 2 is for the same cone as curve 1 except that only the glass sealing lands on the metal shell were enameled to facilitate glass-to-metal sealing; curve 3 is for a tube with a shell made of No. 430 stainless steel. It was because of these and similar data, plus the poor life characteristics of ND14 steel, that the decision was made to stop production of tubes of this construction.

Of the metal tubes, only the type made with a cone of No. 430 steel shows no appreciable increase in pressure on holding. Tubes made with a cone of ND14 metal, however, even after seven weeks, continued to show increase in gas with no indication of leveling off.

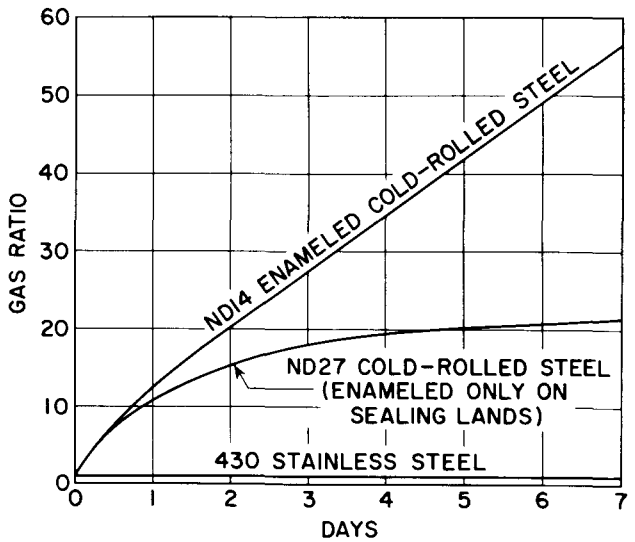


Figure 30. Pressure Stability in Ungettered 21AP4 Picture Tube

Residual Gas Composition

The composition of the gas remaining after exhaust can be as important to tube life as the total gas quantity. Jenkins and Trodden[25] have shown, in their work with impregnated cathodes, that oxygen, water vapor, carbon dioxide, and air have a much greater poisoning effect on emission than do carbon monoxide, nitrogen, argon, and hydrogen.

When changes are made in materials and processing, it is quite possible that the composition of the resulting residual gas can be radically changed with no change in the total quantity of residual gas. For a comprehensive analysis of factors influencing kinescope exhaust, it is, therefore, necessary to have methods of identifying the composition of residual gases. In recent years, several mass spectrometric techniques have been reported for residual gas analysis in kinescopes.[26, 27] At the three Tube Division C & P Laboratories, commercial mass spectrometers have been adapted to residual gas work. Very extensive reports on the study of the outgassing of tube materials have been prepared by Turnbull.[28]

With these methods, it is possible to identify both the type and quantity of residual gases in tubes with known emission performance. By comparing these data with those obtained on tubes processed by any proposed new method, the effect of the new method on life can be fairly accurately predicted.

The data presented in Table V show the residual gas composition for black and white kinescopes. The values reported by Philips and Kemet were obtained with Omegatrons attached to the tubes. The RCA data were obtained with a Consolidated Electrodynamics Model 620 mass spectrometer and an intermediate pumping and collection system to transfer the residual gas from the kinescope to the mass spectrometer. In all cases, the measurements were obtained on ungettered, inoperative tubes. The differences in results probably result from the fact that the tubes were manufactured by different companies using different materials and processing methods. The data, therefore, emphasize that processing and material variations can greatly change the environment of the cathode and can affect reliability.

REFERENCES

1. Code P1
2. Crosslett, V. E., Introduction to Electron Optics, pp. 113-147, Oxford at the Claredon Press, England, 1950
3. Spangenberg, K. R., Vacuum Tubes, pp. 394-411, McGraw-Hill, New York, 1948
4. Moss, H., "The Electron Gun of the Cathode-Ray Tube," Parts I & II; Jour. Brit. IRE, Vol. 6, No. 3, pp. 99-124, January 1945
5. Klemperer, O., Electron Optics, 2nd Ed., University Press, Cambridge, England, 1939
6. Moss, H., "Engineering Methods in the Design of the Cathode-Ray Tube," Jour. Brit. IRE, J. 5, pp. 204-222, Oct.-Dec. 1945
7. Code P2
8. Allard, L. S., "Design Factors in Television Cathode-Ray Tubes," Proc. Inst. Brit. IEE, Vol. 99, Part IIIA, No. 19, pp. 499-507, April-May 1952
9. Maloff, I., and D. W. Epstein, Electron Optics in Television, McGraw-Hill, 1938
10. Moss, H., H. L. Woodbridge, and M. Webb, "Dimensional Tolerances in Cathode-Ray-Tube Guns," Proc. IEE, Vol. 97, July 1950
11. Spangenberg, K. R., Vacuum Tubes, pp. 328-393, McGraw-Hill, New York, 1948
12. Jacob, L., Introduction to Electron Optics, pp. 58-75, Wiley, New York, 1951
13. Zworykin, V. K., and G. A. Morton, Television, 2nd Ed., pp. 159-167, Wiley, New York, 1940
14. Cope, A. D., D. W. Epstein, R. E. Hopkins, S. Lasof, F. H. Nicol, O. H. Schade, and H. Wiedner, "Wide Angle High Definition Television Systems," PB116539, Library of Congress, Washington, D.C., Contract N6onr – 23605, Final Re-

Table V

Source of Data	Per Cent Composition						
	Carbon Dioxide	Carbon Monoxide	Hydrogen	Methane	Nitrogen	Argon	Water
RCA	44	36	19	1	–	–	–
Philips[25]	11	14	17	35	21	.5	1.5
Kemet [26]	–	25	70	–	3	–	–

port, RCA Laboratories, Princeton, New Jersey, August 1952

15. Code P3
16. Baldwin, M. W., Jr., "The Subjective Sharpness of Simulated Television Images," Bell System Tech. Jour., Vol. 19, pp. 563-587, October 1940
17. Leverenz, H. W., An Introduction to Luminescence of Solids, pp. 429-453, Wiley, New York, 1950
18. Soller, T., M. A. Starr, and G. E. Valley, Jr., Cathode-Ray Tube Displays, pp. 609-687, McGraw-Hill, 1948
19. Dushman, S., Scientific Foundations of Vacuum Technique, pp. 757-764, Wiley, New York, 1949
20. Strong, J., Procedures in Experimental Physics, pp. 168-180, Prentice Hall, New Jersey, 1938
21. Holland, L., Vacuum Deposition of Thin Films, Wiley, New York, 1956
22. Bond, W. L., "Notes of Solutions of Problems in Odd Job Vapor Coatings," Jour. of the Optical Society of America, Vol. 44, pp. 429-439, June 1954
23. Hamaker, H.C., H. Bruning, and A. H. W. Aten, "On the Activation of Oxide-Coated Cathodes." Philips Research Rpt., Vol. 2, p. 171, 1947
24. Code P4
25. Jenkins, R. O., and W. G. Trodden, Jour. of Electronics and Control, Vol. 5, p. 393, 1959
26. Wagener, J. S., and P. T. Marth, Jour. Appl. Phys., Vol. 28, No. 9, pp. 1027-1030, September 1957
27. Franchen, J. C., and J. v.d. Wall, International Symposium on Residual Gases in Electron Tubes, Como, September 23-25, 1959
28. Collins, R. H., and J. C. Turnbull, "Evolution and Absorption of Gases in Electron Tubes," Vacuum, Vol. 10, pp. 27-30, February-April 1960

Color Picture Tube Design and Processing

H. C. Moodey and A. M. Morrell

Lancaster

Color picture tubes to be described in this chapter may be defined as directly-viewed cathode-ray tubes capable of reproducing pictures in color. Most color tubes contain screens composed of arrays of three different phosphors which emit light of the three primary colors, red, green, and blue. The phosphors are divided into small areas, arranged so that each picture element comprises an area of each of the three phosphors. The color sub-elements comprising each picture element are small enough not to be resolved by the eye at normal viewing distance.

The color tube must include a means for selecting color so that, as the electron beam scans the screen, only phosphor areas which emit the desired color are bombarded. In some areas, of course, the desired color may be blue, in others, green or red, or a combination of these primaries. Thus, a beam must be directed to a certain phosphor area within each picture element, and directed away or shielded from all other areas of that element.

In addition to the color tubes which contain screens composed of multiple arrays of phosphor elements, several other types have been proposed but not developed. One of these would employ a phosphor, the color of whose output is dependent on beam current density. Lack of phosphors which cover an adequate color gamut is a basic reason for lack of development of this proposal. Another system proposed, known as the "voltage-sensitive," or "penetration," type, uses a screen composed of two or more layers of different phosphors. Varying the beam energy then varies the beam penetration into the screen layers, with consequent shift in the relative excitation of the layers. Although technological advances in applying phosphors have helped reduce the necessary range of beam energy, this range is still several thousand electron volts. Additional problems hindering development include means for obtaining adequate color saturation, raster stability during screen voltage modulation, and suppression of high-frequency radiation.

OPERATIONAL PRINCIPLES OF MAJOR COLOR-TUBE TYPES[1,2]

The color-selection method may be either electrical, mechanical, or a combination of the two. Directly-viewed color tubes may be classified by the color selection method used, as follows: (1) accurate beam scanning; (2) beam switching near the phosphor screen; and (3) direction-sensitive screens.

Accurate Beam Scanning

In concept, the simplest of all color tubes is one in which the screen, composed of horizontal strips of the primary phosphors laid in color sequence, is accurately scanned, strip after strip, by a single beam. Simultaneously, the beam current is modulated by the video signal to provide an intensity appropriate to the color strip and picture element being scanned. A single beam is used, whose focused spot size must be limited to the alloted width of the phosphor line, that is, one-third of the picture element size, in order to prevent color dilution by excitation of an unwanted color. If the screen structure is reasonably fine, then the phosphor strips are narrow, and the practical limitation on light output is the ability of the gun to provide high current in a very small focused spot. The problems of meeting scanning accuracy, and stability requirements at modern scanning frequencies and high definition, make this simple variety of tube impractical.

In its successors, accurate scanning is combined with feedback of a light or secondary-emission signal from the phosphor screen. The signal is generated when the beam (or an auxiliary beam) scans the signal area; it is picked up, amplified, and fed to circuitry which correlates this information with incoming video information. This signal processing provides that modulation of the beam to produce the required output of each primary color occur precisely as the beam scans the corresponding phosphor areas. The signal-generating areas may be located at the end of each scanned line, or may be within the picture area. In the latter case, correcting signals are provided from each picture element or small group of elements. When the signal areas are thus closely spaced, the phosphor screen strips may be horizontal or vertical. In a horizontal-line, single-beam tube, the beam is "wobbled" up and down across a trio of color lines as it scans the picture from side to side. Triple beams may also be used in a horizontal-line tube, in which case each beam is assigned to one color and the three beams may be operated simultaneously. Gun and beam requirements, as well as light output limitations, are similar in kind to those for the simple system without feedback. The feedback tube has been called a "sensing" or "indexing" type. The Philco "Apple"[3,4,5] tube falls in this clas-

sification, as well as several varieties investigated by RCA.[6,7,8,9]

Beam Switching Near the Phosphor Screen

A second type of tube uses beam switching near the phosphor screen. Only one electron beam is used in this type. The switching principle may be subdivided into two categories:

Beam Deflection by a Field Between Elementary Electrodes. An electrode structure, consisting of at least two sets of closely spaced, electrically connected conductors, is supported by the phosphor screen or is placed near it (Fig. 1). When suitable potentials are applied between the sets of conductors, the beam is directed as desired to any one of the three phosphor arrays. Such switching electrodes may have capacitance between sets of conductors sufficiently high to impose a rather severe power requirement for switching color at high frequencies. In addition, objectionable high-frequency radiation may occur from the electrode structure unless proper precautions are taken in its design.

In one form (Fig. 1A) the phosphor lines are laid on, or are coated with, strips of conductive material in color sequence. Adjacent strips are insulated from each other and all strips containing the same phosphor are electrically connected, so that potential differences may be applied between interlaced sets of conductive strips and the beam thus deflected to the desired set. Another form (Fig. 1B), which requires a lower switching voltage, has a structure consisting of sets of small deflecting plates placed near the phosphor screen and registered with the elemental screen areas. In still another version (Fig. 1C) two of the phosphors are applied to the plates and the third to the glass screen plate. The coated screen plate must be sufficiently translucent to pass light emitted by the phosphor coating on the deflecting plates.

A more practical modification of the system of Fig. 1B is shown schematically in Fig. 2. This tube, which has been known by the name of the late Dr. E. O. Lawrence, was developed by Chromatic Television Laboratories as the "Chromatron."[10] Here, the electrode structure consists of a grill of parallel wires. Alternate wires are electrically connected, forming two interlaced coplanar grills. Aligned with the wires is a phosphor screen of parallel strips in a particular color sequence, having portions of each of three color strips opposite each interval between wires. When no potential is applied between grills, the scanning beam strikes the array of phosphor strips located opposite the centers of the spaces between wires; when a suitable potential is applied between the grills, the beam is deflected to strike another array of strips; and when this potential is reversed, the beam is deflected in the opposite direction to strike the third array. Because the wire diameter is small compared with the space between wires, electrostatic focusing at the grill must be used to obtain color purity, a principle which is described under Focus-Grill Systems.

Beam Refraction or Reflection by Accelerating or Retarding Fields. In tubes that use beam refraction or reflection by accelerating or retarding fields (Fig. 3), the electron gun and neck axis are set at an oblique angle to the screen. The beam is shown entering from the

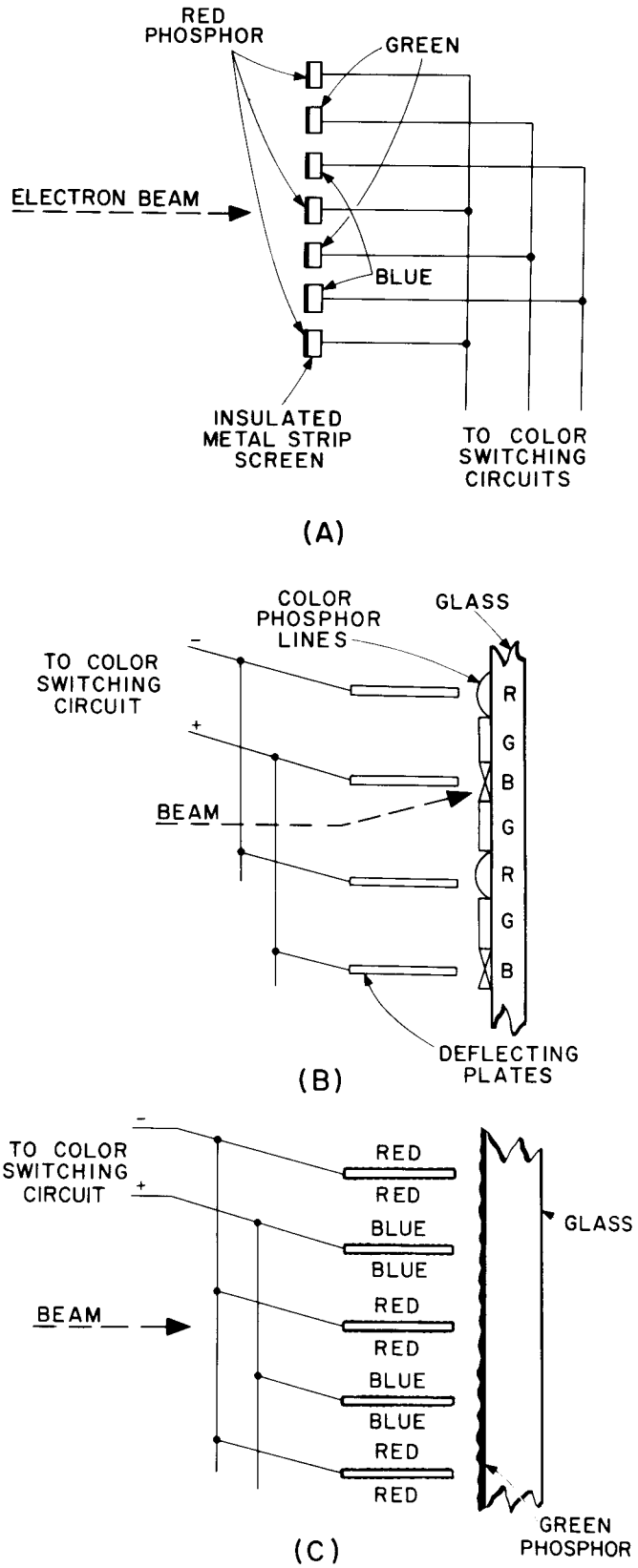


Figure 1. Color Selection by Beam Deflection Between Elementary Electrodes

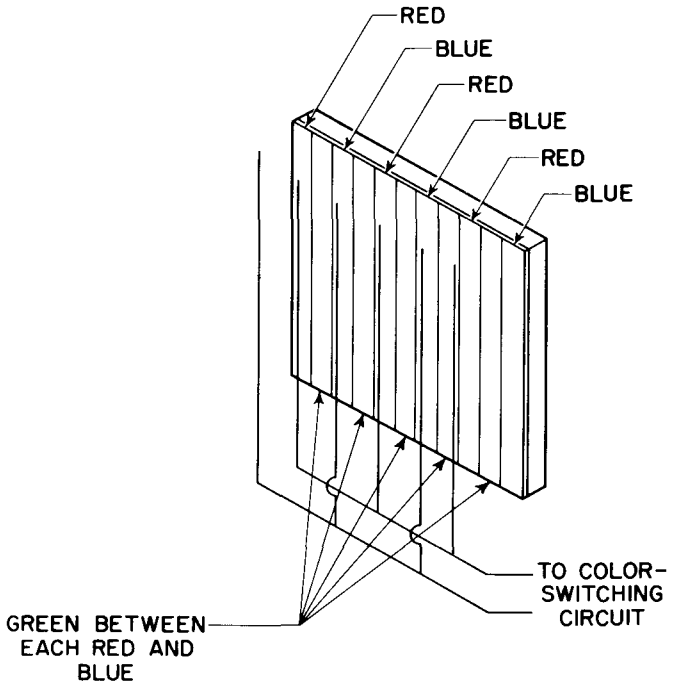


Figure 2. Color Selection by Grill Switching

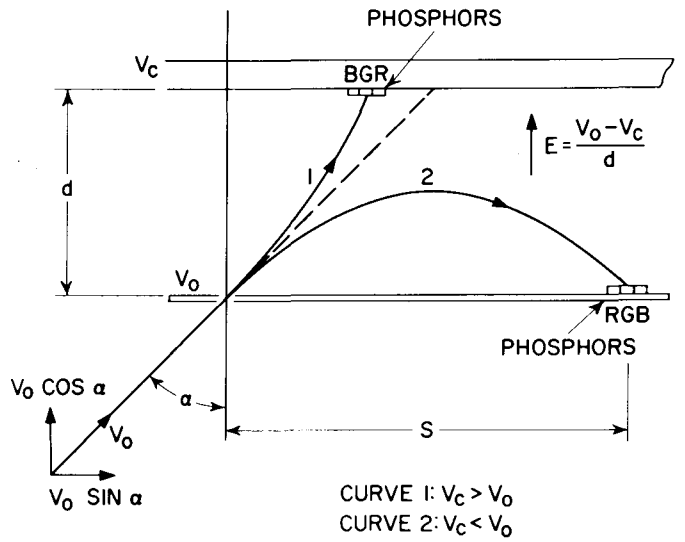


Figure 3. Beam Refraction or Reflection by Accelerating or Retarding Fields

lower left corner of the diagram. The screen structure consists of two electrodes between which a potential difference is applied. These electrodes are an apertured plate, through which the beam passes, and a plate beyond and parallel to it. This figure shows the beam trajectories in the two forms of this tube type.

In the refractive or transmission form (Curve 1) voltage V_c is more positive than voltage V_o and the resultant accelerating field between the two electrodes bends the beam toward the normal, causing it to land on the phosphor areas deposited on the final electrode, the viewing screen. Modulation of the screen potential serves to switch color by changing the refraction of the beam and hence its landing position.

In the reflection form of tube (Curve 2) voltage V_c is negative with respect to voltage V_o and, therefore, the retarding field between the two electrodes reflects the beam which has passed through a slit aperture and causes it to fall on the phosphor-coated aperture-plate surface facing the final electrode. The range S of the beam is a function of the field strength and beam entrance angle α. The phosphor coating consists of a set of three color strips laid in each space between apertures. Because the width of each phosphor strip is about equal to that of the slit, and because focusing action occurs, the beam landing area is narrower than the phosphor strip. Modulation of the reflector electrode voltage provides color switching. The apertured phosphor plate is viewed through the transparent final electrode. Considerable experimental work has been done on the reflection type by RCA[11] and GE.[12]

The focusing action which occurs in both of these forms is shown in Figs. 4 and 5, for a favorable choice of the angle of incidence α and, in the accelerating types, of the voltage ratio V_c/V_o.

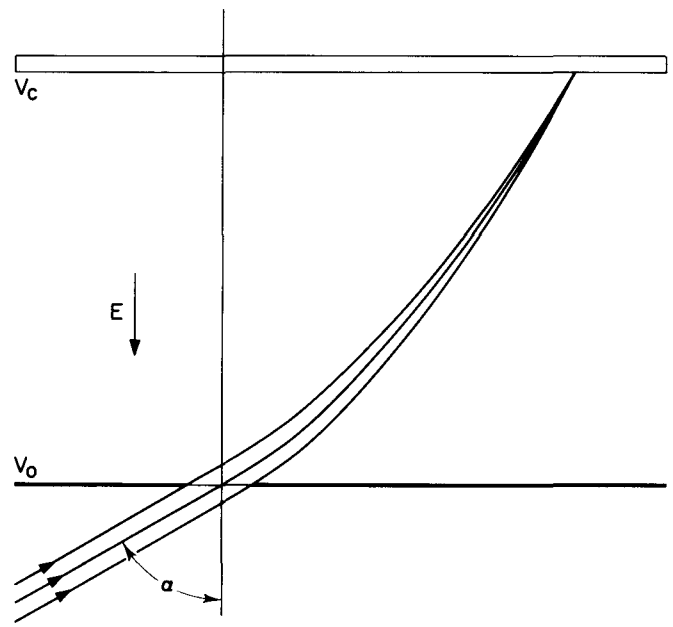


Figure 4. Beam Focusing in Refraction Case

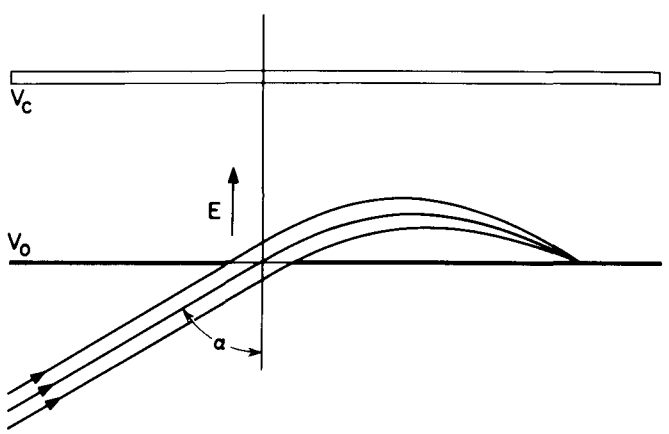


Figure 5. Beam Focusing in Reflection Case

The color tubes in this category require envelopes of unconventional design, with the neck at an oblique angle to the screen. They may be made with somewhat lower capacitance between switching electrodes than the screen-switching types described earlier. The reflection form has the advantage of built-in register between apertures and phosphor arrays; the transmission form, however, may be designed for higher efficiency than the reflection form.

Direction-Sensitive Screen Assembly

Tubes containing direction-sensitive screens require either three beams, or a single beam moved successively into positions corresponding to three beams. (Figs. 6 and 7). A rotating single beam was used in some of the first tubes demonstrated by RCA but was quickly superseded, because of its short duty cycle, by a three-beam arrangement. In the following discussion, wherever three beams are mentioned it should be understood that they can be replaced by such a single-beam system.

Under direction-sensitive screens there are three subdivisions.

Complex Screen Surface Containing Directional Elements. Each of the three phosphors is deposited on surfaces so oriented that only one array of phosphor areas is exposed to bombardment by each electron beam. Several types of configurations for this purpose have been proposed and experimental tubes have been built and demonstrated.[2] Some of these systems (Figs. 8 and 9) use a wide angular separation between electron beams and therefore encounter problems in producing non-keystoned, well-registered rasters; others (as in Fig. 10) use narrower beam angles. It appears that either the angular separation of the beams must be so great as to require separate deflecting yokes, or, for smaller beam angles, a light loss must be tolerated because the phosphors lie on surfaces nearly parallel to the direction of viewing. If separate deflecting yokes are used, severe registry problems arise, requiring precise electrical matching of the yokes and deflection waveforms, as well as accurate keystoning. The basic idea behind such systems is quite old; however, patents[2] have been obtained comparatively recently by Goldsmith and by Geer.

The direction-sensitive screen systems comprising the two remaining subdivisions permit a small angular separation between the three beams. Hence, each employs a single yoke simultaneously deflecting all three beams; this method eliminates the extreme register difficulties just described.

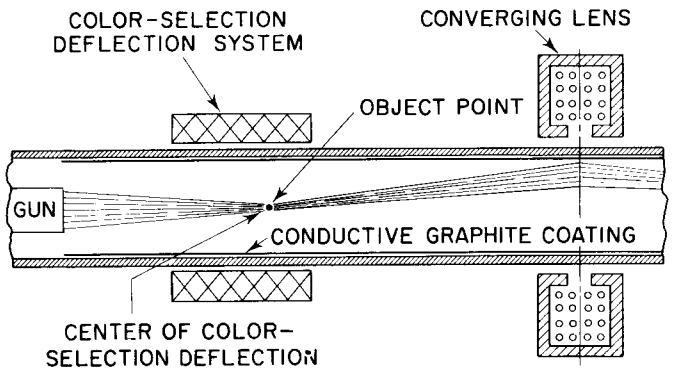


Figure 6. Single-Beam Arrangement for use With Direction-Sensitive Screen (gun end of tube)

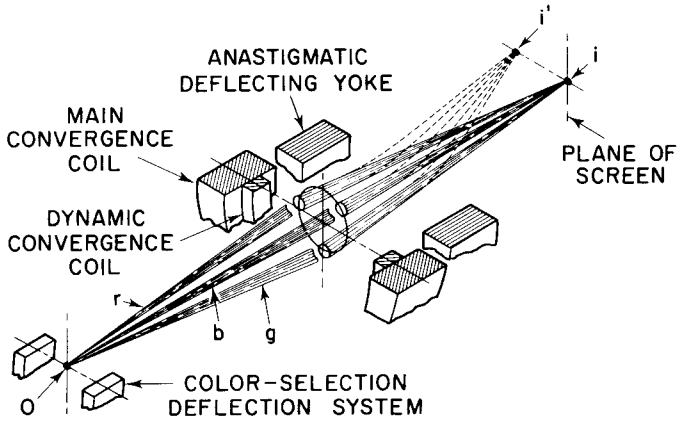


Figure 7. Single-Beam Arrangement for use With Direction-Sensitive Screen (screen end of tube)

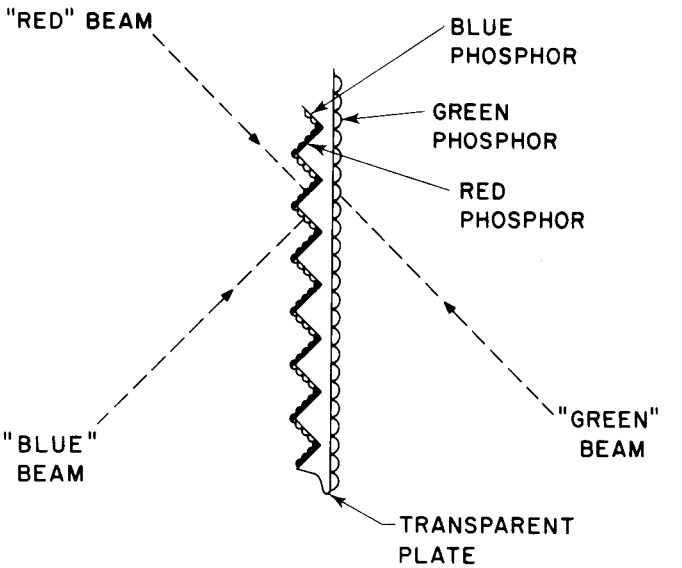


Figure 8. Direction-Sensitive Screen Having Prismatic Surface Elements

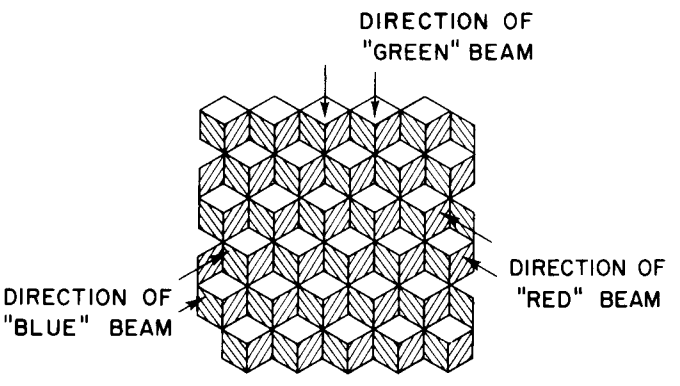


Figure 9. Direction-Sensitive Screen Having Pyramidal Surface Elements

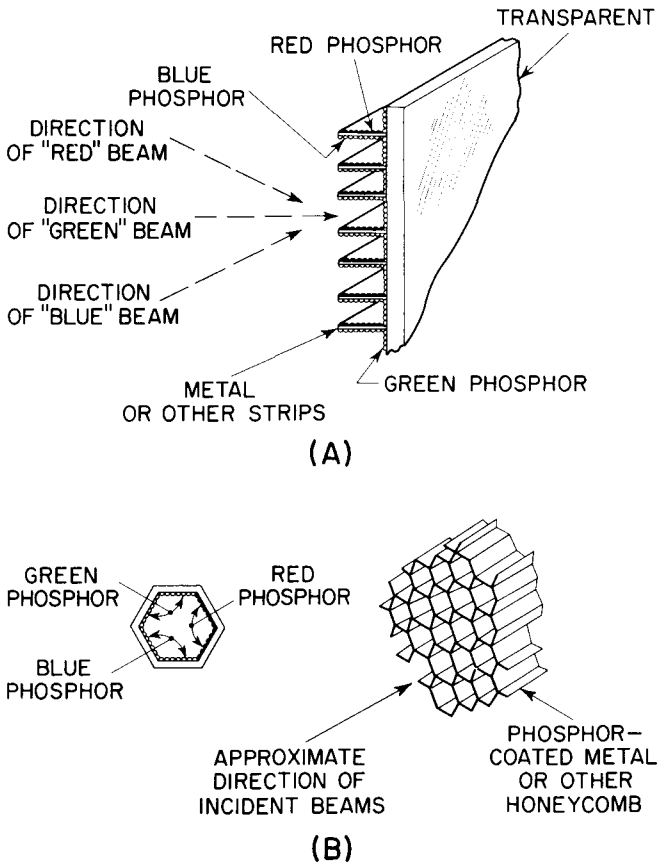


Figure 10. Complex Structures for Direction-Sensitive Screens

Shadow-Mask System.[13] An apertured electrode is placed near the phosphor screen and scanned by three beams, each approaching from one of three apparent sources or "color centers" in the deflection plane of the tube (Fig. 11). Registered with each aperture in the mask is a trio of phosphor dots, one dot for emitting each primary color. The size and location of the apertures and the size and location of the phosphor areas are so chosen that one array of phosphor areas, and only one, can be seen from each color center, looking through the mask (Fig. 12). Because the bulb-wall, mask, and phosphor screen are operated at the same potential, rectilinear propagation of the electrons occurs in this entire region.

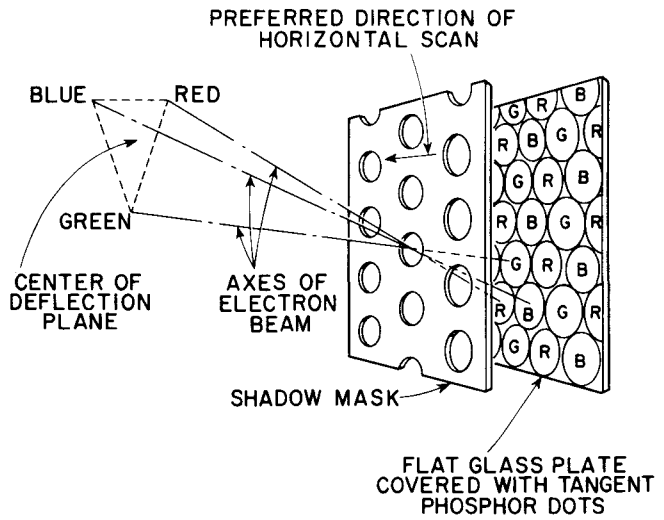


Figure 11. Diagram of Shadow-Mask System

The aperture mask and the phosphor plate may be flat or curved.[14] In the early tubes they were made flat, as a separate assembly mounted within the tube envelope. The mask was placed under tension, so that under normal operating conditions its apertures did not move when the mask was heated by electron bombardment. In the curved form the mask is not stressed, but is so designed and mounted that in normal operation its apertures have very little motion transverse to the beam path. The curved system permits use of a phosphor screen applied to the faceplate of the tube, provided the mask and faceplate curvatures are properly related. Curvature of the assembly reduces the requirement for dynamic convergence and simplifies the problem of obtaining a rectangular (non-pincushioned) raster. Application of the phosphor screen to the faceplate permits full utilization of the bulb diameter, as in black-and-white picture tubes.

The theoretical maximum transmission of any shadow-mask is 33.3 per cent. Specific designs may be more limited, as for instance the round aperture in a hexagonal array, which has a maximum transmission of 30.3 per cent. These values assume a point source in the deflection plane as shown in Fig. 12; they must be reduced when a beam of finite diameter is considered. Also, the maximum permissible hole size is a function of beam separation in the deflection plane, being smaller for close beam spacing. Allowance for these last two factors and a small amount of mechanical error re-

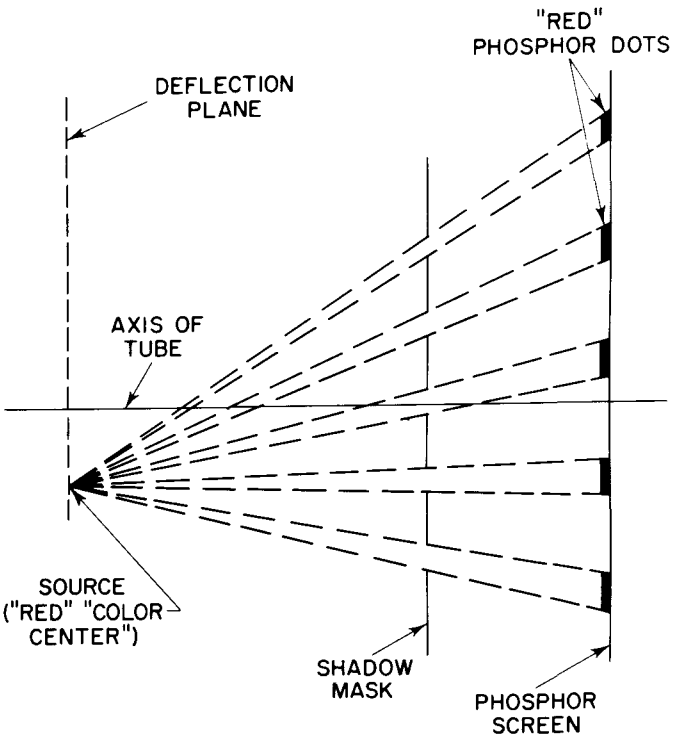


Figure 12. Shadow-Mask Principle of Operation

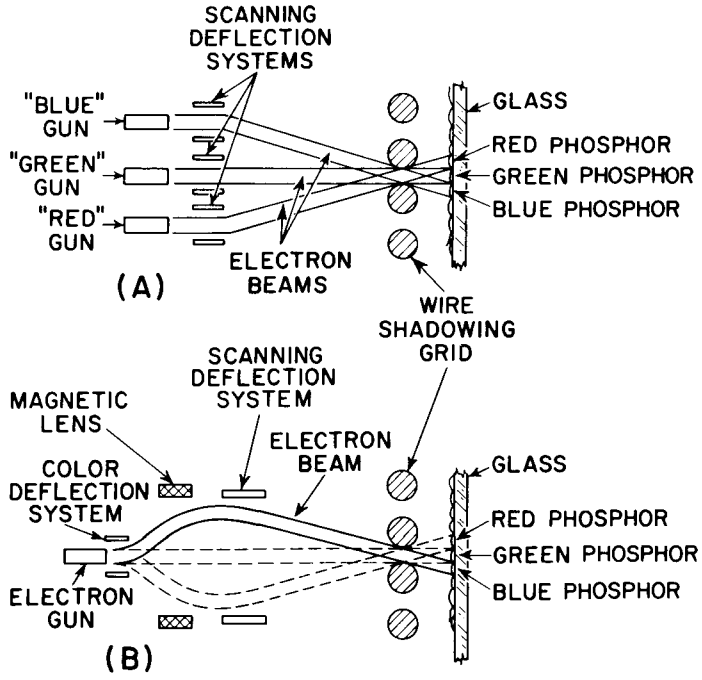


Figure 13. Early Proposal for Line-Structure Shadow-Mask Tubes

quires reduction of mask transmission in a practical tube to less than 20 per cent. Thus, a large fraction of the total beam current is intercepted by the mask, and a corresponding limitation placed upon maximum light output.

The shapes of the apertures and the phosphor areas are optional, but must correspond. Line structures of the shadow-mask type have been described (Fig. 13), but only the type having round apertures and phosphor dots in hexagonal arrays has been produced commercially.

The so-called "thin" tube of Dr. D. Gabor [15, 16] also uses the shadow-mask principle for color selection.

Focus-Mask[17] and Focus-Grill[18,19] Systems. As in the shadow-mask type, an apertured electrode is placed near the phosphor screen, aligned with corresponding screen areas, and scanned by a beam approaching from each of the three color centers. In focus-mask or focus-grill types, however, the apertures are relatively large. An accelerating field on the screen side of the apertured electrode produces an electron lens at each aperture. These lenses collect the electrons in each beam, which would otherwise bombard the three phosphor arrays indiscriminately, into small pencils which strike only one array (Fig. 14).

The focusing electrode may be of high transmission (up to about 90 per cent for the grill structure) so that it intercepts relatively little beam current and consequently removes the severe limitation on light output imposed by the shadow-mask.

The accelerating field between focusing electrode and screen results in loss of contrast due to two types of electron scattering. In the first type, secondary electrons emitted from the focusing electrode (upon beam bombardment) are accelerated to the screen; in the second type, primary electrons, back-scattered from the screen coating at some loss in velocity, are re-accelerated to the screen. Focusing systems using an additional, or auxiliary, grill or mesh to prevent this loss of contrast have been studied and investigated experimentally, especially by RCA.[19] However, the additional transmission loss, complexity of construction, and problems connected with placing the screen on the curved faceplate, have combined to make this form of tube unattractive.

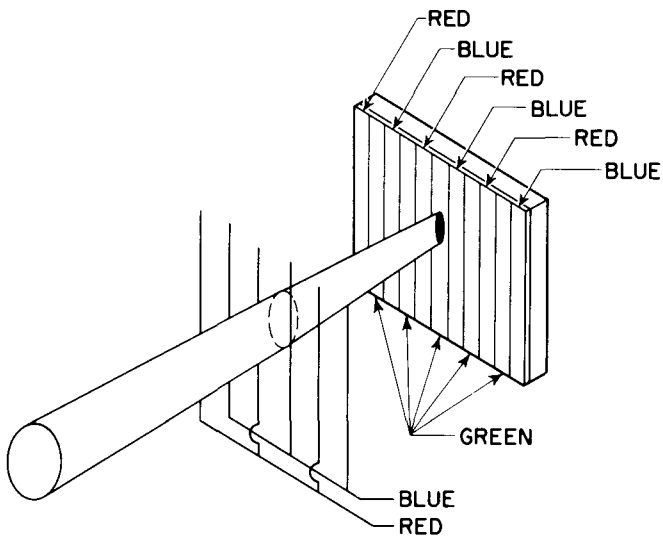


Figure 14. Focus-Grill Operation

Because of the fact that the beams are focused after deflection, these types are sometimes called "post-deflection focus" tubes.

As shown in Fig. 14, the focus-grill principle may also be used with a single gun. In this case, color selection is obtained by grill switching. This combination was mentioned previously in connection with the Lawrence tube.

THE GEOMETRY OF SHADOW-MASK TUBE DESIGN

Planar Mask-Screen Assembly[13]

The basic design dimensions of a shadow-mask tube are determined by straightforward geometry. When the mask and screen are curved, the geometry becomes a little more complex than in the case of a flat screen assembly because of the additional variables introduced, namely, the curvatures of face and mask.

In Fig. 15 a flat mask and phosphor plate are shown parallel to each other, spaced at distances p and L, respectively, from the source, which is a distance S from the axis of the assembly. This sketch shows only one of the beams; the other two would be represented by similar diagrams rotated 120 degrees around the axis. The center-to-center spacing of the apertures in the mask is a, and the center-to-center spacing of the phosphor dots of one color is D. In the phosphor-screen application process*, and later during tube operation, the mask pattern is projected onto the phosphor plate, forming an image having dimensions D/a times the corresponding mask dimensions. This ratio D/a is equal to L/p, which is termed the magnification λ. The magnification in this planar structure is constant, regardless of the angle at which the rays strike the mask. Consequently, if the mask-hole spacing is uniform, as indicated by the dimension a, the phosphor-dot spacing D in each array is likewise uniform. By the same reasoning, the dot size and the dot-to-dot spacing within each trio are also uniform, and each trio of dots is a replica of every other trio.

Fig. 16 shows the basic design parameters for a planar screen assembly. The algebraic derivation of the equations relating these parameters is quite simple.[13] A basic objective is optimum utilization of the screen area, which requires the best possible packing arrangement for the phosphor dots (tangent dots in hexagonal array) as indicated at the right of the diagram. Best packing also implies constancy of dot, and dot-trio, size and spacing. Since practical mask fabrication requires uniformity of aperture spacing, uniform trio spacing can be obtained only when the screen magnification is constant over the entire picture area. As already mentioned, this constancy is inherent in the planar assembly.

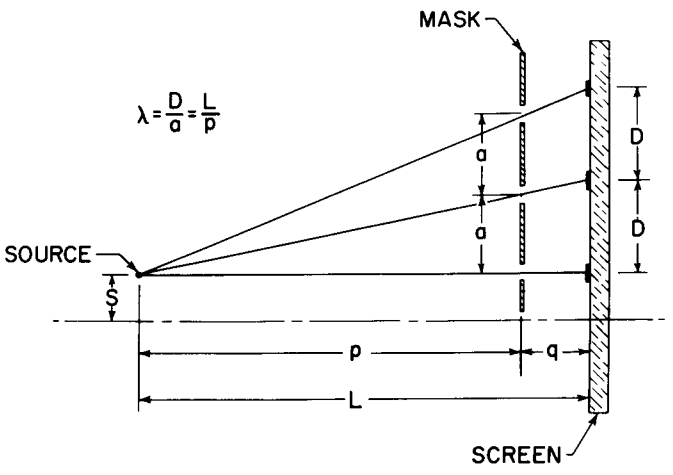


Figure 15. Planar Mask-Screen Geometry

Note that Fig. 16 takes into account the finite diameter M of the electron beam in the deflection plane. If the aperture diameter is given by B', electrons will bombard the entire phosphor dot. However, if any slight error in register of mask with screen occurs, electrons will strike adjacent phosphor dots. Practical considerations therefore require the introduction of a "screen tolerance" by reducing the aperture diameter somewhat to the diameter B given in the section of this chapter on DESIGN EQUATIONS. The bombarded area of each phosphor dot is thus decreased to provide a margin of safety between the fluorescent area and the edge of the phosphor dot as shown in Fig. 17. Thus, to ensure pure color fields, mask transmission is sacrificed to some degree.

From these considerations, it becomes apparent that the "color centers," shown as points in Figs. 11 and 12, should be redefined as small areas in the deflection plane. All beam electrons passing through one of these areas, and only those electrons, are directed so that, unless intercepted by the mask, they will strike the phosphor dots of one array. In the case of zero screen tolerance, the beam of diameter M exactly fills the color center, as shown in Fig. 16.

Several relationships developed from the geometry of Fig. 16 (see section on DESIGN EQUATIONS) have important effects on tube design. First, for a given cone length and mask-hole spacing, the mask-to-screen spacing q varies inversely with S, the beam-to-axis spacing in the deflection plane. Second, either increasing the beam diameter M in the deflecting plane or decreasing the value of S reduces the maximum permissible aperture diameter B and hence the transmission of the mask. However, enlargement of S to increase transmission aggravates the problems of beam convergence, which are discussed further in several sections of this chapter. Upper limits for the phosphor-dot diameter and the aperture spacing are set by the degree of screen coarseness which is permissible.

Curved Mask-Screen Assembly[22,23,24]

Thus far, only the case of a flat mask and screen assembly, a type of construction used within the envelope of the first production tube, the RCA 15GP22, has been considered. For these tubes, the phosphors were applied by a "silk screening" process. However, with the development of photographic techniques for phosphor deposition, the application of screens directly to the spherical faceplate of the tube became feasible. For several reasons, both mechanical and electrical, the masks used with these screens were likewise spherically curved.

* "Lighthouse" design features are discussed in Refs. 20 and 21.

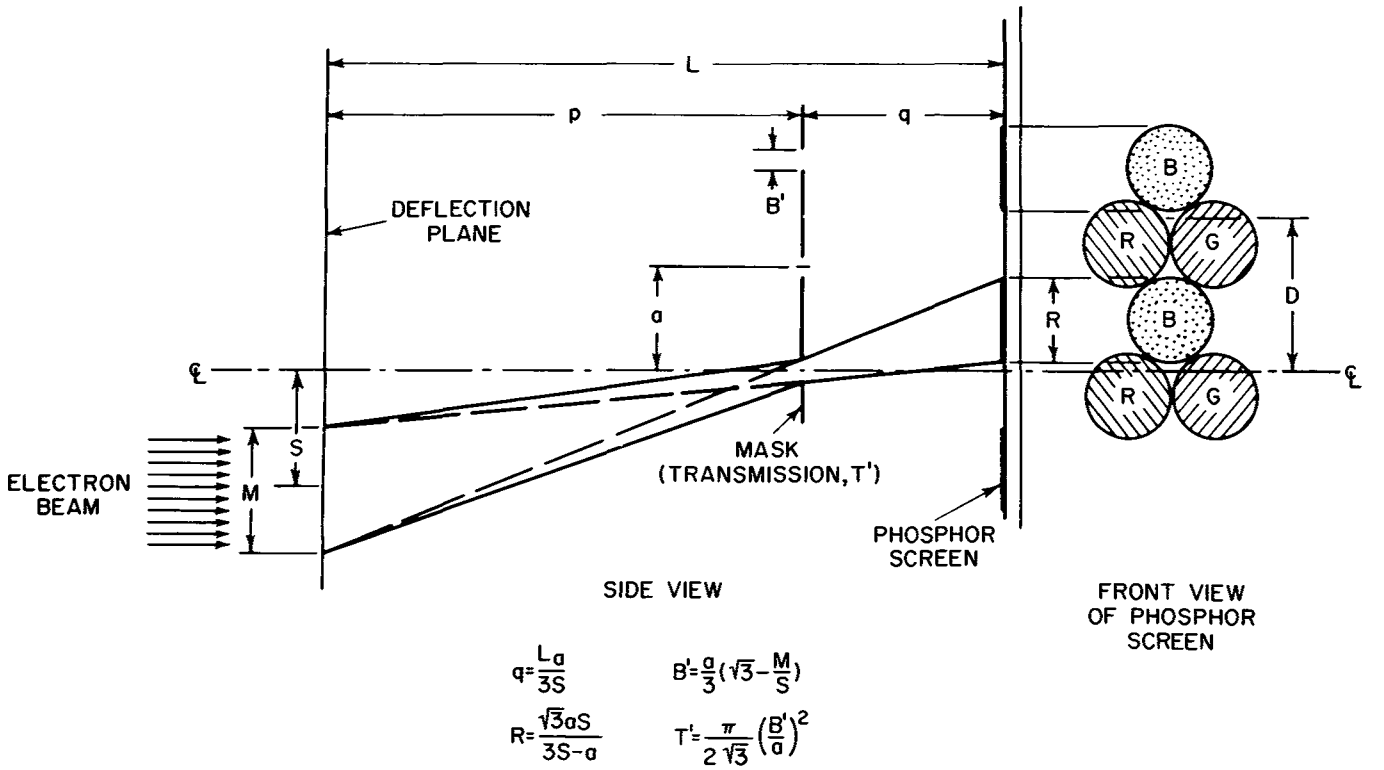


Figure 16. Shadow-Mask Design Parameters

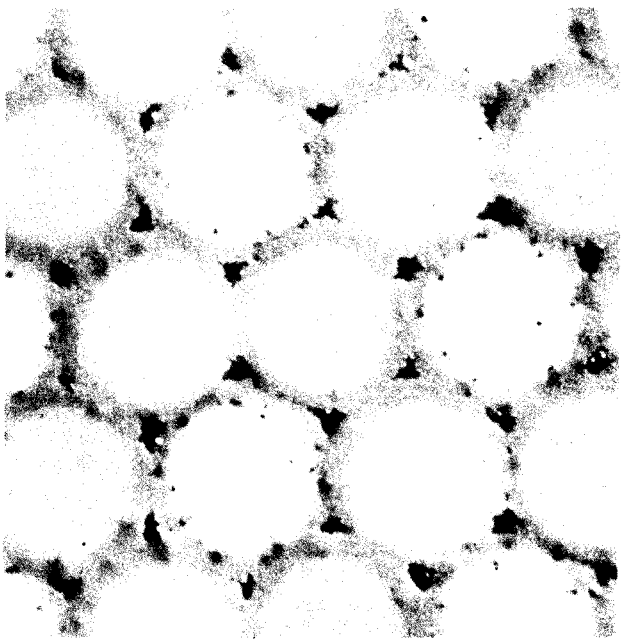


Figure 17. Magnified Portion of Screen of Operating Tube, Showing Bright Fluorescent Areas Within Phosphor Dots

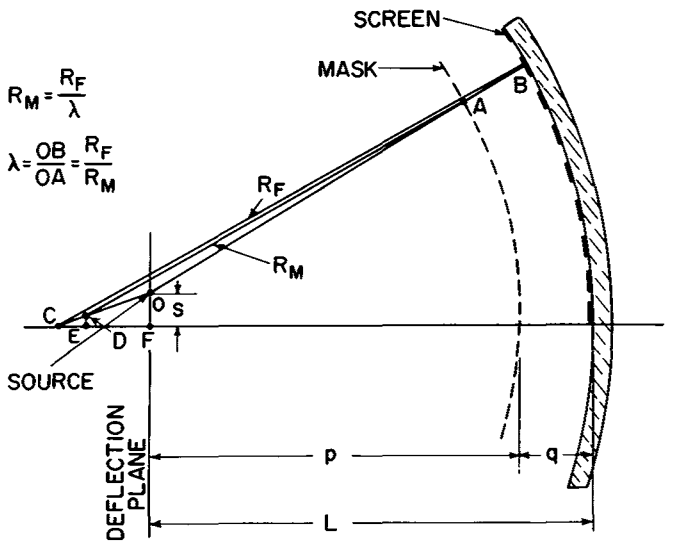


Figure 18. Curved Mask-Screen Geometry

Fig. 18 shows the additional factors which must be considered in the design of a tube with a spherical-screen assembly. The geometry of this tube is worked out by first assuming a suitable radius of curvature for the faceplate and then determining the coordinate mask curvature needed to provide constant magnification. The appearance of the picture demands that the faceplate be of rather long radius of curvature; consequently, its center lies at point C, back of the deflection plane OF. The geometry of the system can then be worked out to a first approximation for the condition of constant magnification, that is, for a constant ratio of lengths OB to OA at all deflection angles, where B is any point on the screen. A slight compromise is needed to take into account all three beams, that is, the center of mask curvature must be shifted from point D to point E on the axis, an adjustment which has negligible effect on the constancy of magnification.

The geometry of the spherical system, in contrast to that of the planar system, is such as to produce asym-

metrical instead of equilateral trios toward the edge of the screen.

This trio distortion, a quasi-optical effect termed "crowding," may also be derived geometrically. The object, which is the set of three color centers in the deflection plane, is broken into two components, one in the plane within which the beams are deflected, and the other perpendicular to this plane. Magnification ratios are then derived for and applied to these two components separately, to obtain the corresponding dimensional components of the image on the screen, which is the trio of phosphor dots. Fig. 19 is a diagram in which the tube axis and the rays corresponding to deflected beams lie in the plane of the paper. The object size in this plane is w, and the corresponding image dimension (as read through a microscope having its axis parallel to the tube axis) is y cos β. If x is constructed parallel to y, and constant screen magnification λ is assumed, then

$$\frac{y}{x} = \frac{L - p}{p} = \lambda - 1$$

In the small triangle constructed at the object,

$$\frac{x}{\sin (90 - \Theta_0)} = \frac{w}{\cos (\Theta_0 - \beta)}$$

and, since $\Theta_0 \cong \Theta$ for practical cases,

$$x \cong \frac{w \cos \Theta}{\cos (\Theta - \beta)}$$

The magnification ratio relating the object in the deflection plane to its image on the curved screen is then

$$m = \frac{y \cos \beta}{w} = \frac{x}{w} (\lambda - 1) \cos \beta = \frac{\lambda - 1}{1 + \tan \Theta \tan \beta}$$

Thus, the radial phosphor-dot separation is a function of radial position of the dot trio on the screen, even though λ is constant. Values of this function (also called the "obliquity factor") depend on the relative magnitudes of L and the radius of curvature of the screen. The plane perpendicular to that of Fig. 19, containing the deflected ray OP but not the tube axis, is considered next (Fig. 20). Obviously, no such foreshortening of the object takes place because w_n, the object projected onto this plane, lies parallel to the image y_n. Thus, the magnification of this component is

$$m_n = \frac{y_n}{w_n} = \lambda - 1$$

Fig. 21 shows the composite effect on trios, as viewed from the front of the tube. This diagram is not to scale, but is exaggerated for clarity. The overlapping of the phosphor dots may be lessened somewhat by an optical compromise at the time of screen exposure, described in the following section.

ELECTRON-OPTICAL DEVIATIONS FROM THE SIMPLE OPTICAL ANALOGY FOR SHADOW-MASK TUBES

To a first approximation, simple geometrical principles apply to the electron-beam behavior and hence to the photographic method of screen deposition for shadow-mask tubes.[20,21] The screen exposure is made in a "lighthouse" within which, as shown in Fig. 12, light is projected from a source through the mask apertures to the screen plate. The source must be of

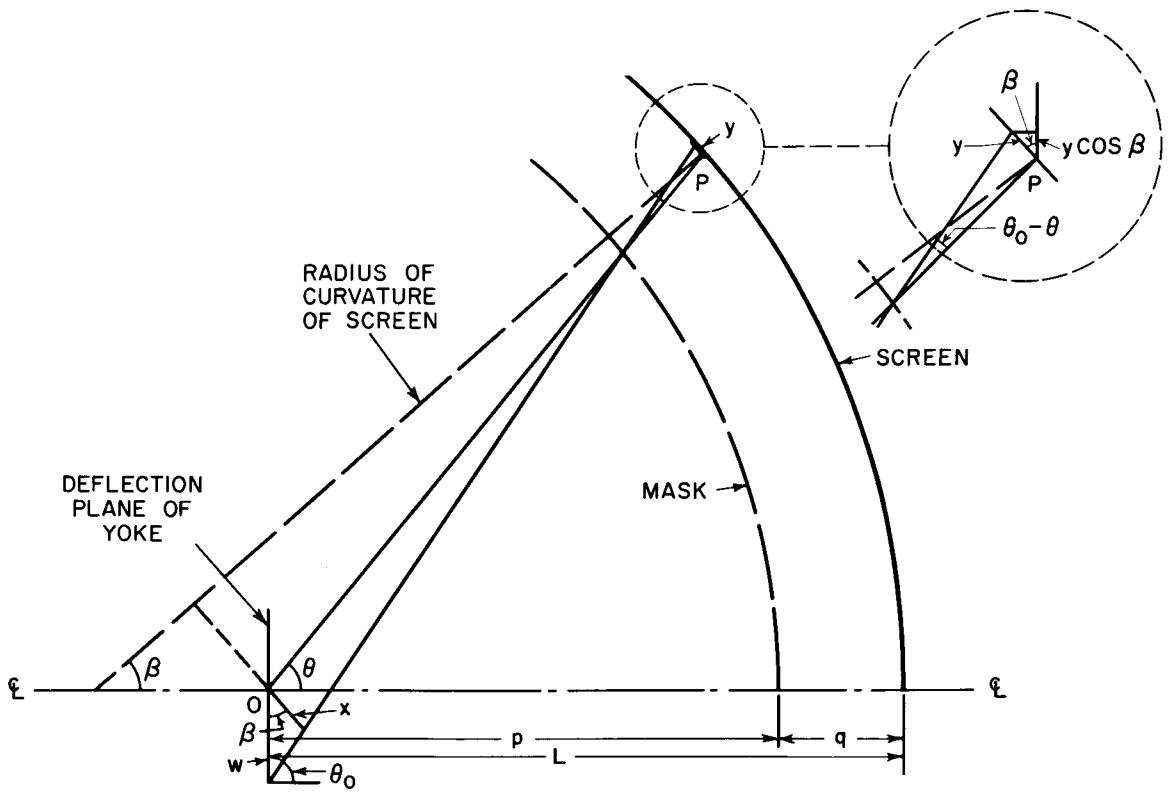


Figure 19. Geometry of Trio Distortions (plane containing tube axis)

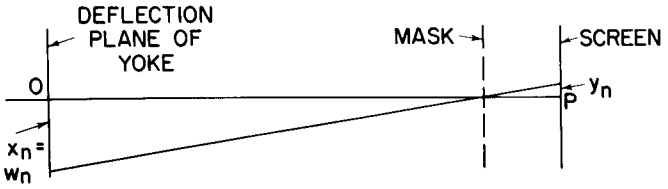


Figure 20. Geometry of Trio Distortions (plane normal to Fig 19)

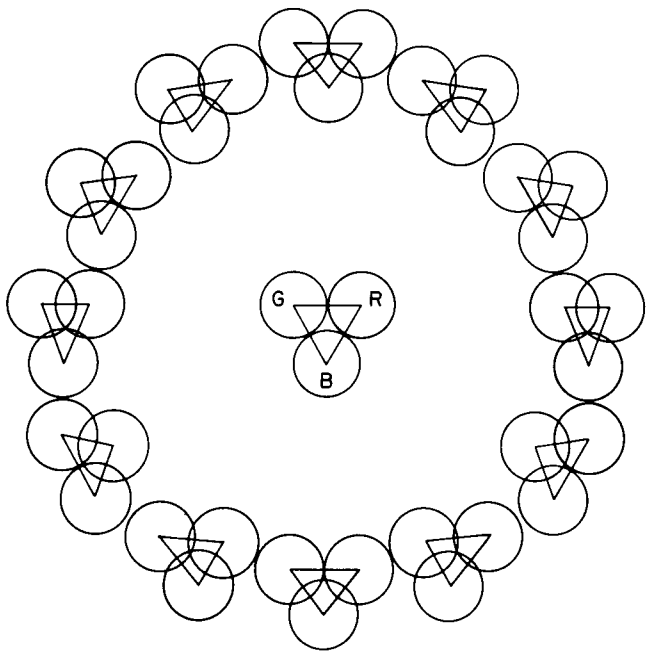


Figure 21. "Crowding" of Trios on Tube Screen

small size and suitable shape, and must be properly located. The screen plate is coated with a photosensitive phosphor mixture.

As development of the shadow-mask tube progressed in the direction of wider deflection angles and higher-quality screens, it became apparent that several electron-optical phenomena were preventing the shadowed beams from landing accurately centered on the corresponding phosphor dots. This misregister caused excitation of unwanted colors. Pure color fields could, of course, have been obtained by designing for a large screen-tolerance but only at considerable reduction of aperture size and light output. A consistent effort has therefore been made to build into the tube corrections for all types of systematic misregister, allowing sufficient screen-tolerance only for random and other errors not readily corrected.

Correction for the systematic errors is made in the lighthouse at the time of screen exposure by simulating optically the electron-optical effects which will occur later in the operating tube. These misregister effects include: (1) radial error, (2) degrouping errors, and (3) errors due to any ambient magnetic field.

Radial Error

A radial error results from the fact that the effective beam-deflection center moves toward the screen as deflection is increased (Fig. 22). This characteristic

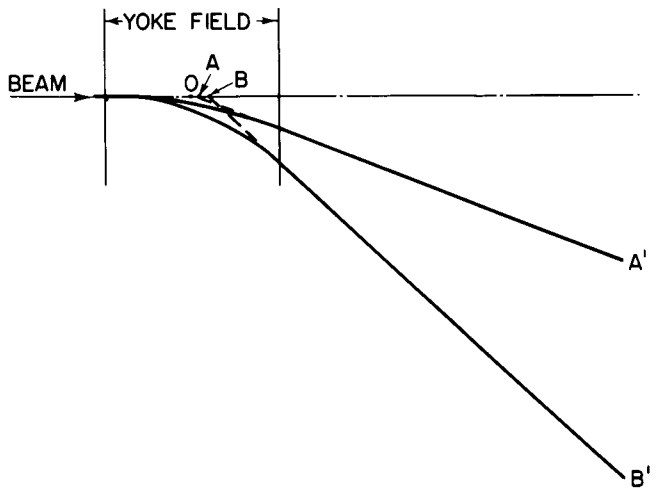


Figure 22. Forward Motion of Deflection Center With Increasing Deflection Angle

is a function not only of the tube parameters but also of the yoke length and field distribution. The manufacturing variations among yokes made according to a particular design do not, however, affect the characteristic substantially. Hence, a radially symmetrical lens, designed to fit the characteristic of the tube-plus-yoke system, effectively corrects this type of misregister.[22,23,25] Curves 1, 2, and 3 of Fig. 23 represent the characteristics for three yoke positions, in order of increasing screen-to-yoke distance.

To obtain the lens design[26] it is necessary first to convert the data displayed in Fig. 23 into a relationship between two other variables: the virtual source position (analogous to the moving deflection center position in the operating tube), and the corresponding points at which rays from the virtual source intersect the lens surface. These derived data are then fed into a computer which has been programmed to furnish the required lens contour or "grinding instructions."[27]

Various considerations govern the detailed procedure for carrying out these steps. The misregister characteristic (Fig. 23) is generally determined by averaging measurements on several tubes, taking as misregister for each screen radius the average of a number of azimuthally spaced readings. The yoke used during these measurements must be of the same design as that which will be used to deflect the lens-corrected tube*. As indicated by the differences between the three curves of Fig. 23, the characteristic is critically affected by the yoke (i.e., deflection center) position with respect to the screen. In general, it is desirable to choose a lens design for which a minimum amount of glass is removed in fabrication, and in which the slope of the surface is kept to a minimum. Such a lens will not only require less grinding but will also tend to be more accurate than a deeply cut lens. Furthermore, its placement in the lighthouse will be less critical. Curve 2, which roughly balances radial misregister on the plus and minus sides of zero, corresponds to the correction which an optimized lens will furnish. From any one of

* See also under Yoke

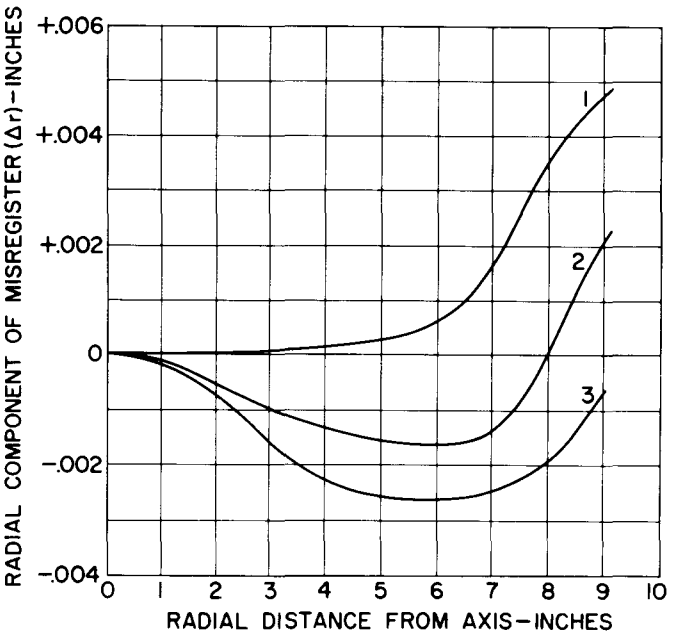


Figure 23. Radial Misregister Characteristic of 70° Shadow-Mask Tube

the characteristics, a set of points as required by the computer program can be selected for conversion into the new variables.

When a radial correction lens is used, a minor optical adjustment is generally made to reduce the geometrical trio distortion ("crowding") shown in Fig. 21. If the lens is offset outward (away from the tube axis) from a position centered over the light source, and is rotated with the source successively into the three exposure positions, the radial lens characteristic is such as to reduce the "crowding" of the phosphor dots and thus provide more nearly uniform screen coverage. The effect of this offset is shown in Fig. 24. Here the source is a distance S to the left of the tube axis, and the lens center is offset a distance B to the left of the source. The line joining the source and the screen center indicates that optically the lens center is displaced a distance A. Diametrically opposite rays, striking the screen at equal distances r from the tube axis, intersect the lens at unequal distances from its center ($\rho' < \rho''$). Since the refraction produced by the lens increases from its center outward, the ray reaching point D lies at a greater angle to the tube axis than the ray reaching C. Thus, the dot formed at D will have been displaced outward from the axis by more than the average radial correction for screen radius r, and the dot formed at C will have received less than the average outward correction. If this figure is assumed to represent the condition during exposure of the "blue" dot array, point C corresponds to the "blue" dot at the top of the screen and point D to the "blue" dot at the bottom of the screen. Referring to Fig. 21, then, the lens offset will move both of these dots downward from the positions shown and thus reduce dot overlap or "crowding." The "blue" dots to the right and left of the screen center will not be moved by the offset, but exposures for "red" and "green" arrays will move the corresponding dots apart.

Degrouping Errors

Degrouping errors result to a large extent from the motion of the beams relative to each other produced by dynamic convergence fields. These ac fields are provided in synchronism with scanning, to maintain the beams in convergence over the entire picture area.* The continual change in the value of S due to this beam motion causes each beam, projected through a mask aperture, to change correspondingly its landing positions relative to the phosphor dots. The entire "front end" of the tube, from deflection plane to screen, may be considered as an imperfect example of a multiple-pinhole camera in which the mask apertures are the pinholes. As the screen is scanned from center to edge, the object in the deflection plane, which is the trio of beam cross-sections, is enlarged by dynamic convergence so that its image, the trio of fluorescent spots on the screen, is correspondingly enlarged. But if the phosphor dots were deposited on the faceplate by exposure from a fixed light source (constant value of S), the beam trio is magnified or "degrouped" with respect to the corresponding dot trio, as shown in the left column of enlargements (Fig. 25).

Considering one beam at a time, the degrouping is a unidirectional vector throughout the picture area; the magnitude of this vector is a function of beam position on the screen. This magnitude, averaged around a zone of a given radius, increases with screen radius in proportion to the square of the tangent of the deflection angle measured from the tube axis. Around each zone, there is an azimuthal variation of two cycles (Fig. 26), of which the peak-to-peak amplitude increases with screen radius. Furthermore, the azimuthal characteristics for the three beams are separated in phase. This separation is not 120 degrees, as might have been expected because both the phasing and the wave shape are affected by the dynamic convergence which is, in turn, related to the rectangular raster shape. The magnitude and phase differences between the degrouping characteristics of the three beams distort the shapes of the triangles formed by the beam trios so that the symmetrical "crowded" triangles (corresponding to the dot trios shown in Fig. 21) are no longer observed.

As an early expedient to reduce the effects of degrouping on picture quality, a compromise adjustment was made in the direction of grouping the beams, thus reducing the screen tolerance at the center to obtain more tolerance near the edge (right column of enlargements in Fig. 25).[23] This adjustment, used in the 21AXP22A and known as a "short q" or "compromise S" arrangement, consisted of setting the mask-screen spacing for an "S" value intermediate between the zero-deflection and full-deflection dynamic conditions, and exposing the screen from this intermediate "S" position.

A much better solution to the degrouping problem has been incorporated in the 21CYP22.[24,28,29,30] The motion of each beam in the deflection plane, which is equivalent to a transverse shift of the source, is simu-

* Further discussion of convergence will be found in several sections of this chapter.

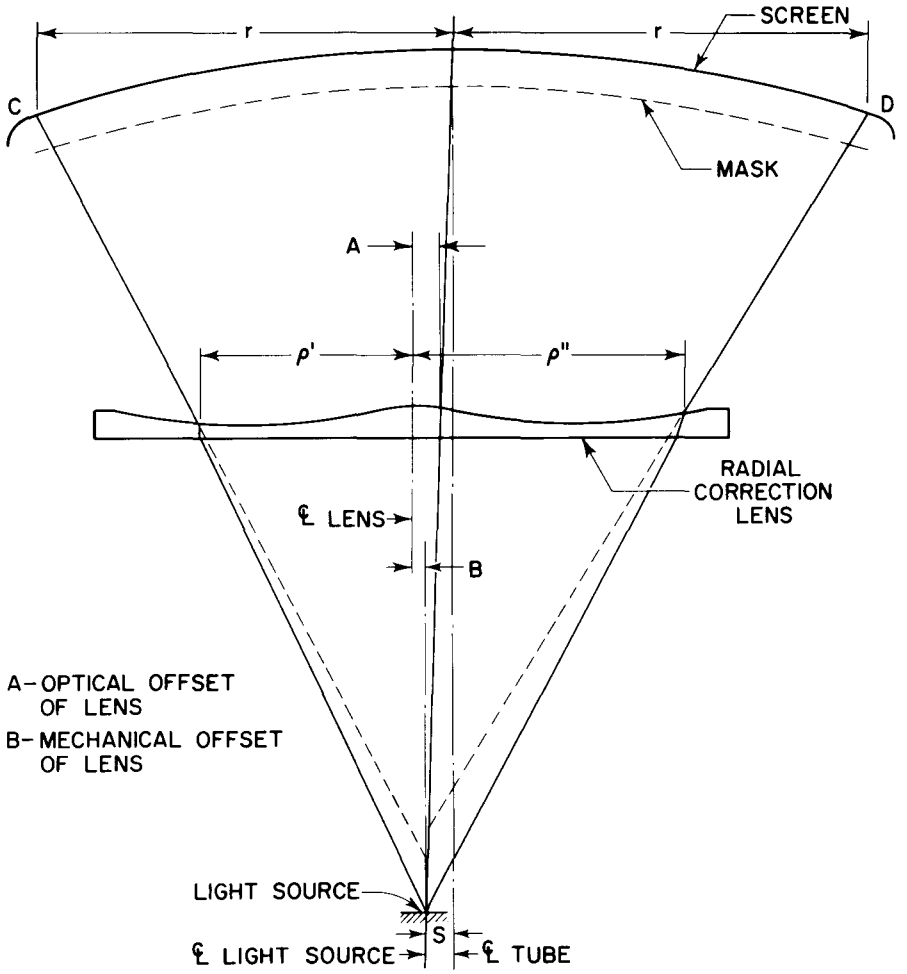


Figure 24. Diagram of "Crowding" Reduction by Use of Lens Offset

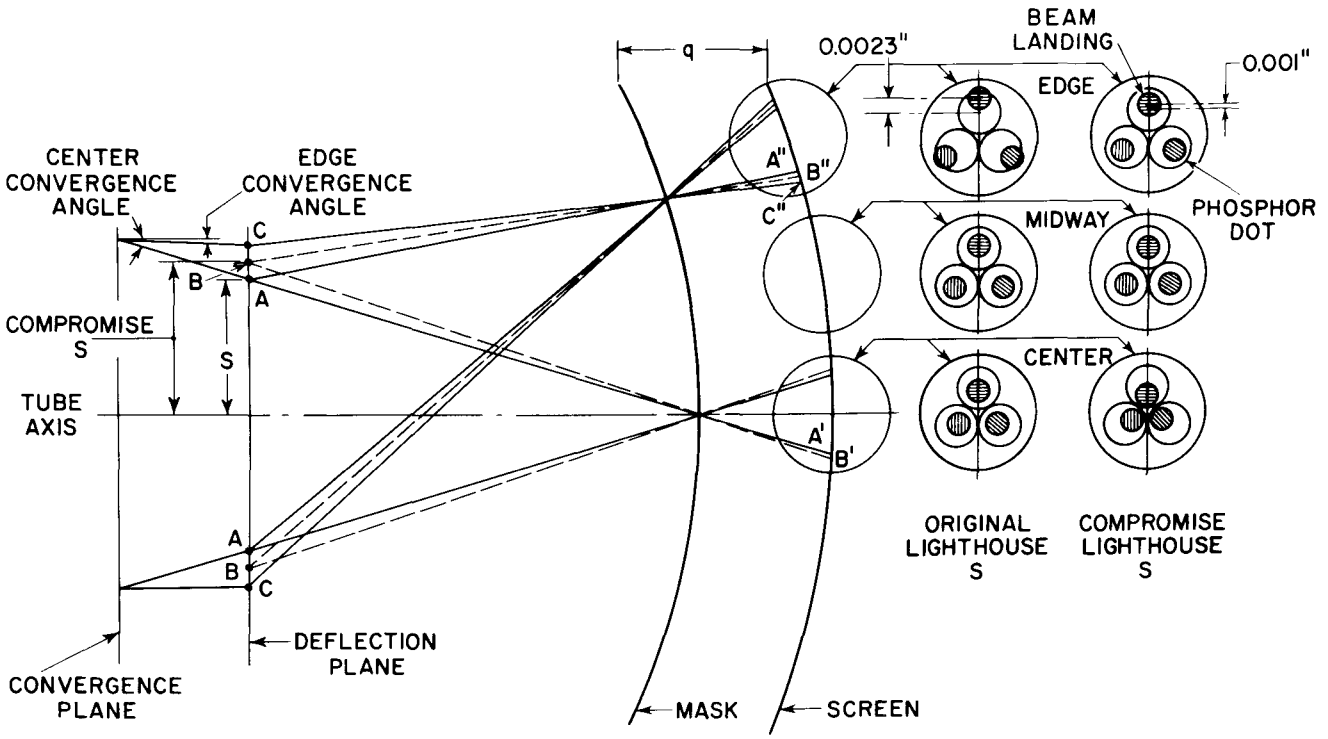


Figure 25. Effect of Dynamic Convergence on Register

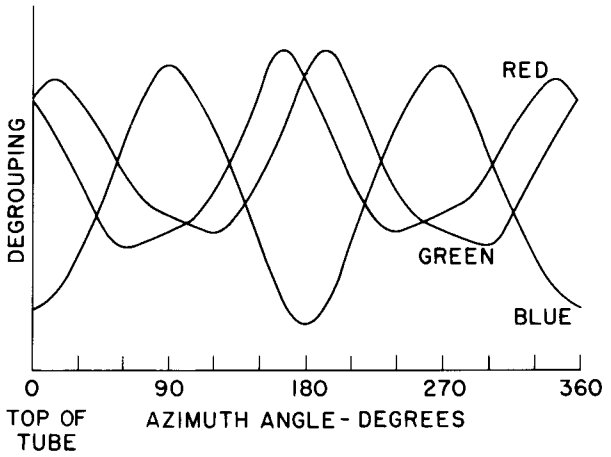


Figure 26. Azimuthal Variation of Dynamic Degrouping

lated optically by the introduction of a prism. This prism is not of the simple type having two planar, non-parallel surfaces; although the tilt around any elemental zone of this special prism is constant, the slope increases continuously from center to edge of the lens. Because this lens is mounted in fixed relationship to the source in the lighthouse, and this assembly is moved successively into positions representing the three color centers for exposure of the three phosphor-dot arrays, this optical system can simulate the motions of the three sources, averaged for each zone and phased 120 degrees apart by use of the three lighthouse source positions. However, the azimuthal variations around a given zone displayed by any one beam cannot be reproduced by the uniform prism around the corresponding lens zone. In fact, this zonal uniformity of prism does not even correct for azimuthally uniform degrouping; it produces an azimuthal characteristic of its own known as "printing errors." Unfortunately, in shadow-mask tubes these printing errors are not so phased as to compensate for the azimuthal variations in degrouping, but they do tend to reduce the dot "crowding" effect.

Thus far, attention has been directed toward means for laying phosphor dots in the locations where the degrouped electron beams would strike the faceplate. However, in a "constant magnification" tube, these dot trios would be degrouped and the dots of adjacent trios would overlap, thus reducing the useful phosphor area. The remedy for this condition is a redesign of mask contour in the direction of decreased curvature, so that mask-screen spacing, and hence the magnification ratio*, are gradually reduced from center to edge. The mask contour is so designed as to provide in each zone the exact grouping effect needed to nullify the average degrouping. This correction is effective, of course, for both light and electron beams.

In practice, the radial and degrouping corrections are combined in a single lens surface (Fig. 27). No mechanical offset between this lens and the light source is needed for "crowding" correction; the slight optical

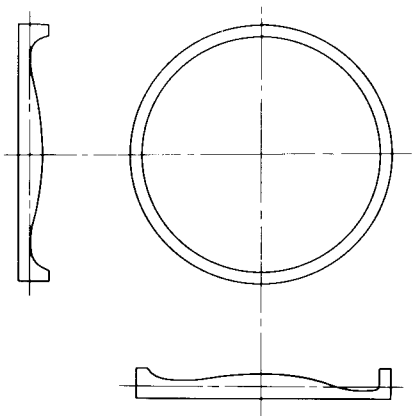


Figure 27. Contour of Degrouping Lens

offset inherent in this arrangement (mentioned previously under Radial Error) is sufficient.

Use of the degrouping correction lens and redesigned mask contour in the 21CYP22 removes the need (exemplified in the 21AXP22A) for grouped beams at the screen center. The regained tolerance in the central portion of the screen may then be sacrificed to raise light output by enlarging the mask apertures in this area. In the 21CYP22 it has thus been feasible to grade the mask-aperture diameter upwards from 10 mils at the edge to 12 mils at the center providing a 44 per cent increase in mask transmission at the screen center as compared with that of the 21AXP22A.

Although the degrouping lens provides a major correction in shadow-mask tubes, a small error remains as previously mentioned; furthermore, the amount of this incorrigible error increases rapidly with deflection angle, so that extension of these principles to tubes having appreciably wider deflection angles is considered impractical.

Ambient Magnetic Fields

Ambient magnetic fields such as that of the earth cause register errors.[23,31] Partial compensation for the component of such a field transverse to the tube axis may be made by offsetting the mask-screen assembly in the lighthouse in a direction perpendicular to the transverse field. As the orientation of the tube axis in the horizontal plane is random, and man-made fields, if present, may have any direction, the only field to which all tubes are consistently subjected is the vertical component of the earth's field. Consequently, the lighthouse offset has been applied only to correct for a weighted average magnitude of this vertical field in the United States. Further discussion of magnetic field effects will be found later in this chapter.

TRIPLE-BEAM GUNS

Only the design and operational features peculiar to multi-beam guns will be considered here, since basic gun design for cathode-ray tubes is covered in other chapters.

* For this "pinhole camera," the magnification ratio is the ratio of the mask-screen spacing to the mask-deflection-plane distance.

Special Requirements

Triple-beam guns for shadow-mask color kinescopes have several special requirements, in particular: (1) their beams must converge everywhere in the picture area; and (2) they must deliver higher currents than guns in most other cathode-ray tubes, to compensate for the inefficiencies of the mask-screen assembly.

Convergence. Convergence may be obtained mechanically (by tilting the individual guns toward their common axis), or electron-optically. Fixed mechanical tilting provides convergence in only one zone of the screen. Other screen areas require different angles of tilt because the beam path-length between deflection plane and screen varies with deflection, and also because the deflection field itself has some converging action. When center convergence only is applied, the degree of off-center misconvergence is roughly proportional to the beam spacing in the deflection plane and to the square of the tangent of the deflection angle.

In practice, convergence of the beams is maintained over the picture area by dynamic electronic control operating in synchronism with the scanning. Several means for applying dynamic convergence have been investigated. At RCA, the first method developed was the use of an electrostatic accelerating lens at the top of the gun, common to the three beams.[32,33] Varying the potential of one of the electrodes which formed this lens provided the necessary change in beam convergence. In the 15GP22 gun assembly, which used this principle, the individual guns were parallel to each other and to the axis, and the final lens then provided both the static and the dynamic convergence needed. Two of the major shortcomings of this gun were the requirement for close regulation of the convergence lens potentials, and the lack of individual beam control inherent in the common lens.

The succeeding gun, which is used in current color tubes[22] (Fig. 28), overcame both of these difficulties. Through three pairs of internal radial pole pieces at the top of the gun, a magnetic convergence field is applied to each beam and varied in strength as needed, with little effect on the other two beams (Fig. 29). For added flexibility in obtaining accurate convergence, a tangential pair of pole pieces is added to the "blue" gun, to obtain a "lateral" or tangential dc adjustment of this beam (Fig. 30). In addition, mechanical convergence is used in this gun, thus bringing the required dc convergence fields close to zero and obviating requirements for extremely good regulation of power supplies.

Suitable dynamic fields can be supplied to the radial pole pieces to provide excellent convergence all over the picture area, but this essential correction gives rise to the previously described "degrouping" effect (Fig. 25). Because the transverse center plane of the radial pole pieces (called the "convergence plane") lies about four inches back of the deflection plane, the application of convergence fields between these poles causes the separation of the beams in the deflection plane to vary. If the convergence plane could be moved forward to coincide with the deflection plane, degrouping would

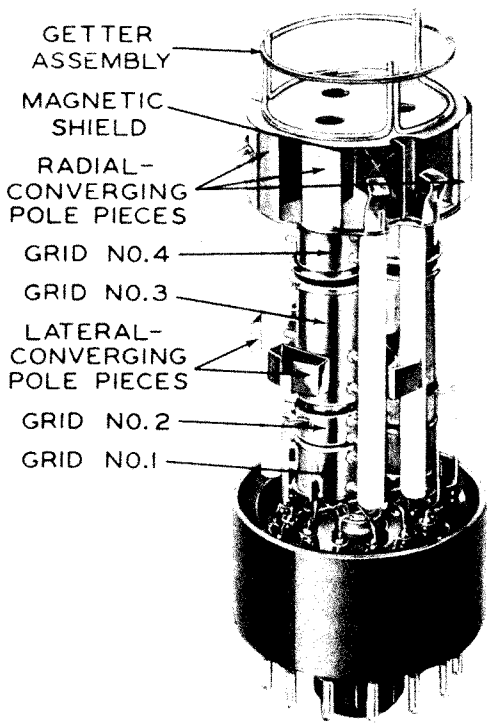


Figure 28. Photograph of 21CYP22 Electron Gun

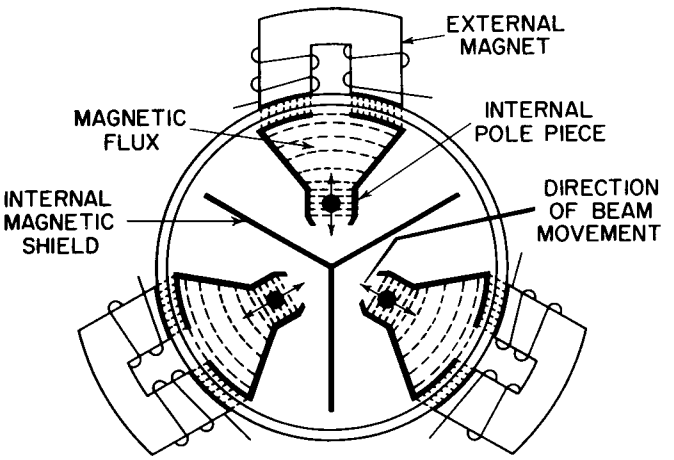


Figure 29. Schematic of Radial-Converging Pole Pieces

vanish. The proximity of these two planes is limited in practice, however, by distortions due to interaction of the components and fields which provide deflection and convergence.

High Beam Current. Although color picture tubes are operated at higher voltages than other types, the efficiency of the mask-screen assembly is so low that considerably higher beam currents are required as well. The "red" gun, in particular, has to operate at high beam current because of the relatively low efficiency of the "red" phosphor. The use of high cut-off voltages, together with correspondingly high video drive, produces a focused-spot size. However, this approach is limited by the effect on life of the higher cathode loading imposed by such operation.

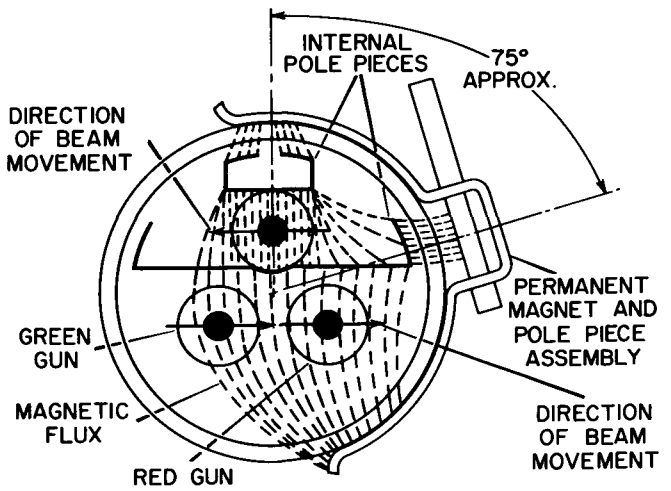


Figure 30. Schematic of Lateral Converging Pole Pieces

Refinements have been made in present standard guns to produce reasonably good focus and very small beam diameter in the deflection plane. This latter characteristic is important in relation to mask transmission and screen tolerance, as mentioned in the section on Geometry.

Description of Gun Types

The individual guns of the 21CYP22 gun assembly (Fig. 28) have electrically isolated cathodes, control grids, and No. 2 grids. Acceleration-type focus lenses are formed between grids No. 3 and grids No. 4, which have common connections for the three guns. Separate connections to all of the lower gun elements are needed to meet drive requirements, as explained later under Drive Characteristics.

In the course of developing the present gun and tube, many variations in gun design have been investigated. In addition to the standard triple-cylinder design, a triple-aperture type[32,34] has received considerable attention. In the triple-aperture type (Fig. 31), each electrode is common to all three guns, and consists basically of a disc containing a set of three punched apertures. This "horizontal" construction has an advantage over the standard "vertical" design in that no transverse relative motion of the gun parts can occur. The spacings between the three guns can therefore be made very accurate and stable. However, this construction does not lend itself readily to electrical separation of control grids and grids No. 2. A hybrid design may also be used, in which the lower electrodes are separately fabricated and electrically isolated, while the upper parts, forming the focusing lenses, consist of a pair of triple-aperture discs. A ceramic-aluminum gun using this principle has been made.

Other gun investigations have included the so-called "in line" type.[34] In contrast to the standard "delta" arrangement, in which the individual guns surround the tube axis at 120 degree separation, the beams in an "in line" gun lie in a horizontal plane with the center beam on the tube axis. It is easy to see that such an arrangement will function in place of a standard gun

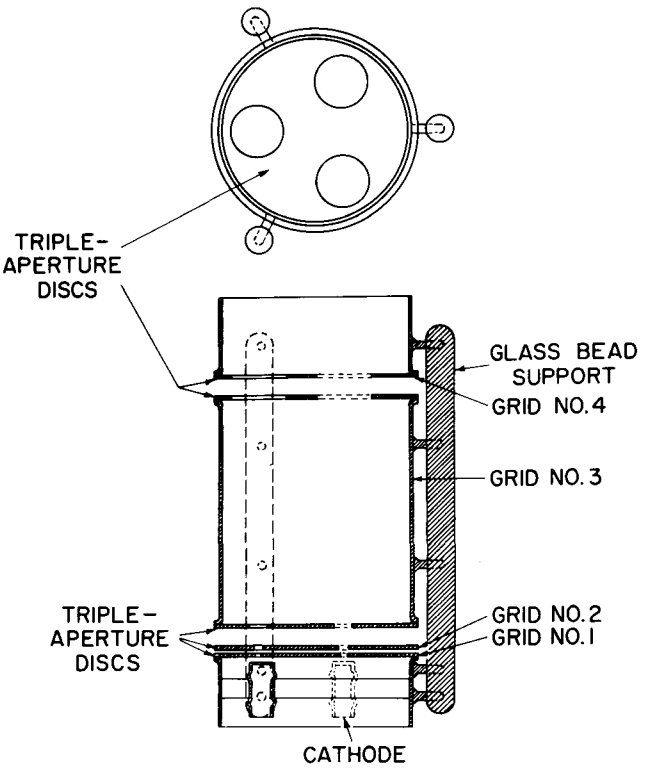


Figure 31. Schematic of Triple-Aperture Gun

when one realizes that multiple sets of "color centers" exist in the deflection plane (Fig. 32). By definition the beams may pass through any set of red, green, and blue color centers to produce fields of pure colors. A "first order" and a "second order" color center are designated by 1 and 2 on this diagram. Location of higher-order color centers may be worked out in similar fashion, resulting in an array of color centers shown in Fig. 33. Here, the standard, first-order centers are B_1, G_1, and R_1. However, if an "in line" gun is to be used, the centers chosen are G_2, B_1, and R_2, shifted downward to place B_1 on the axis. Strictly speaking, these higher-order color centers are well defined only for a planar mask-screen structure; in the curved version they are smaller, out of round, with hazy boundaries. For this, and for convergence reasons, use of the higher-order color centers is not attractive; furthermore, going to higher-order color cen-

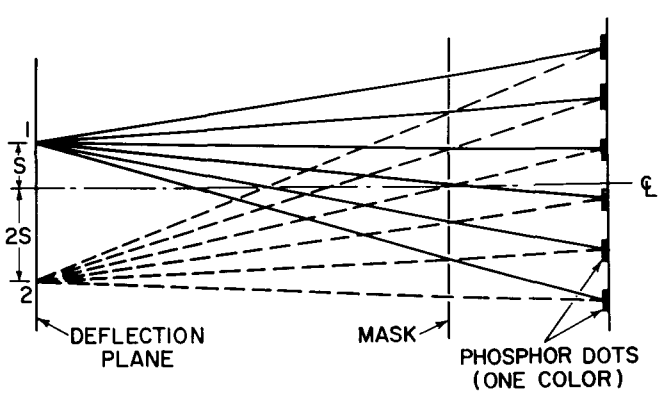


Figure 32. First and Second Order Color Centers

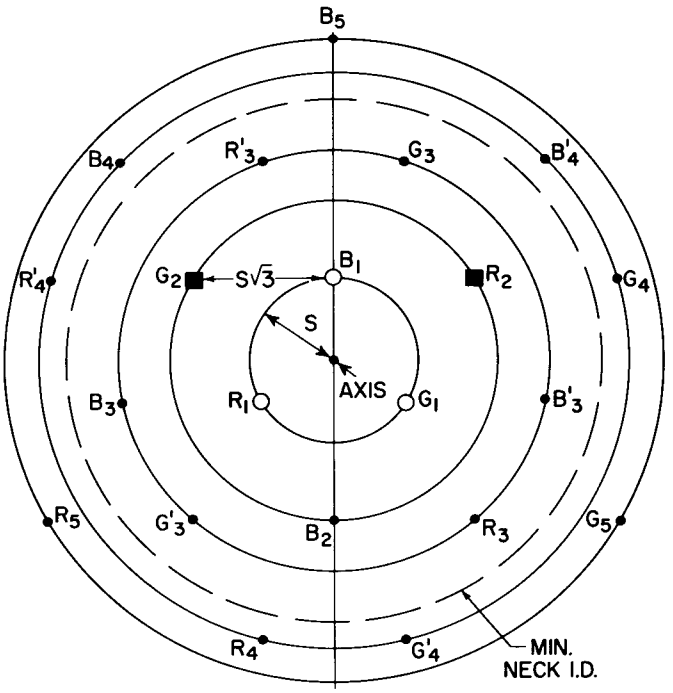


Figure 33. Array of Color Centers

ters does not permit enlargement of apertures as is generally the case with increased "S." Other considerations are that this gun is less compact and that it shows more extreme azimuthal variations in degrouping than does the "delta" gun.

MECHANICAL DESIGN FEATURES OF SHADOW-MASK TUBES

The geometry of the shadow-mask tube presents critical mechanical restrictions on tube construction. These restrictions have led to unique construction of internal parts of the tube in addition to special features in the bulb.

The mechanical features of the metal tube, the 21AXP22 and the 21AXP22A, and the glass tube, the 21CYP22, are covered in this discussion.

The Shadow-Mask-Frame Assembly

The apertures in the shadow-mask are formed by a photoengraving technique which permits accurate control of aperture size and spacing. The mask is etched in a flat strip and later formed to its approximately spherical contour. The forming operation stretches the mask slightly, but with production tooling this stretch is consistent and, hence, may be compensated for in the original printing master.

Early masks were made of copper-nickel alloy, 0.008-inch thick, but with the 21CYP22, 0.006-inch thick cold-rolled steel was used. The etching and forming problems initially associated with steel have been solved and steel's greater strength has minimized mask denting and damage during tube fabrication. The lower thermal expansion of steel minimizes expansion problems in tube processing and operation, and the lower material cost represents a substantial saving. The magnetic properties of the steel mask are also desirable.

The skirt of the formed mask is welded to a cold-rolled-steel frame. Fig. 34 shows the details of one of the three springs which are welded to the frame and engage tapered studs secured to the panel sidewall. This system of three-point support allows differential expansion of the panel-and-mask-frame assembly during tube fabrication. The flexing of the springs absorbs the movement during thermal cycles and allows the mask to return to its original position upon cooling.

This mounting system meets three basic requirements:

1. Replaceability of the mask within the top cap with an error of less than 0.0005 inch. Such accuracy is required because the mask must be removed a number of times during the screening operation and, therefore, any error in replacement position would directly affect relative phosphor dot placement between the three phosphor fields.
2. Stability throughout processing. As discussed previously, the system must allow for differential expansion of parts and the return of the parts to their original position.
3. Mechanical strength. Any assembly must be mechanically strong enough to withstand handling during tube fabrication. In addition, the mounting in the finished tube must be rigid enough to pass drop and vibration tests.

The frame used is made of cold-rolled steel fabricated from strip material, wrapped and welded. To minimize movement through tube-processing cycles, the frame is stabilized by a 500 C, three-hour cycle while it is supported from the three points corresponding to the spring attachment area.

In addition to the three requirements mentioned above, the expansion of the mask due to electron bombardment during tube operation must be minimized. Because the

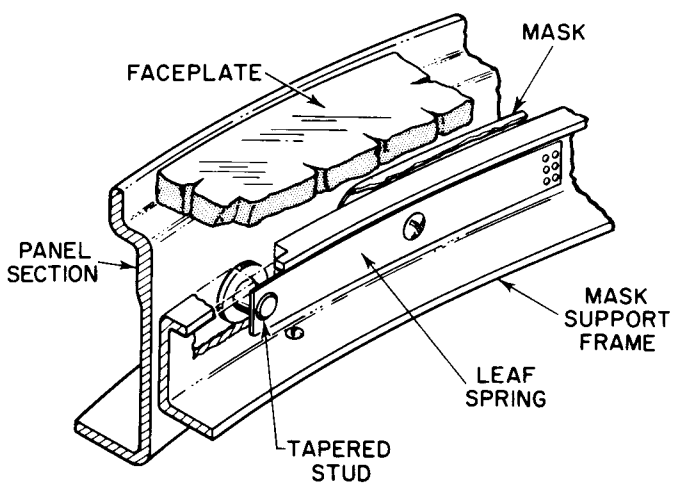


Figure 34. Details of Spring Mounting of the Shadow-Mask Support Frame (as used in the 21AXP22A)

mask transmits only about 15 per cent of the electron beam current, and typical receiver supplies deliver about 20 watts, appreciable power must be dissipated by the mask.

Increasing the emissivity of the mask and frame minimizes the heating and expansion of the mask. The emissivity of the funnel coating is also an important factor. Fig. 35 shows the radial misregister caused by thermal expansion with various combinations of mask and funnel coatings.

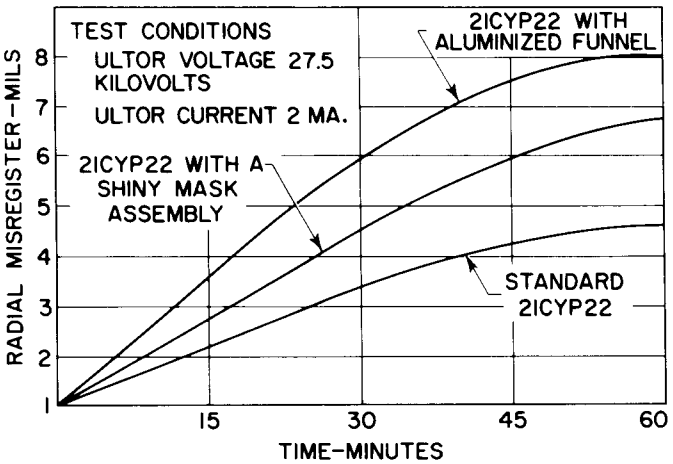


Figure 35. Relative Radial Misregister Due to Mask Expansion Caused by Beam Bombardment

The steel mask used in the 21CYP22 is steam blackened to give high emissivity as well as to prevent rusting during tube fabrication. Iron plating and chemical blackening was used with copper-nickel masks but was not satisfactory in the glass tube because of the high temperature air bake required for bulb sealing. This high temperature air bake caused the mask coating to peel.

Because of the large mass of the mask and frame, the expansion of the mask and frame does not follow rapid variation in ultor current. From a cold start, 30 minutes to an hour is required for the mask to reach its final temperature and expansion for any given ultor power. This time lag is shown in Fig. 35. The misregister produced by mask expansion is radial and, hence, can be corrected by repositioning the yoke. A compromise position of the yoke, made after the tube has warmed up, often gives the best purity.

The design of the mask-frame assembly, must provide for minimization of the scatter of stray electrons due to overscan. These electrons could reduce contrast and color saturation. In the frame design, the lower leg of the "L" effectively keeps the electron beam from scanning the inner vertical surface of the frame and bouncing back to the screen. An "electron shield" of thin steel sections (Fig. 36) is welded to the frame at final mask insertion to reduce the space between the frame and the panel sidewall. A small clearance is maintained to accommodate differential expansion. Minimizing this gap prevents electrons from "sneaking" around the frame assembly and bouncing off the panel sidewall into the screen.

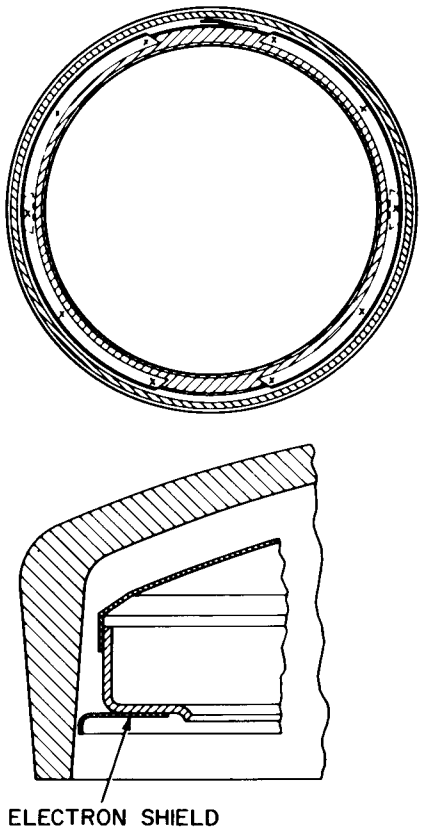


Figure 36. Electron-Masking Shield

The design of the mask apertures must take into consideration the possibility of "strays" (scattered high-velocity and secondary electrons) generated at the mask. The beam does not strike the mask normal to its surface, except in the center, and, because the mask thickness is comparable to the aperture diameter, a tapered or conically shaped aperture must be used to prevent the beam from striking the mask-aperture sidewalls. Fig. 37 shows this effect. The taper is not uniform in this mask but increases toward the edge as the angle of the beam increases. Tapering the aperture also increases the transmission upwards of 20 per cent at the mask edge because the beam is not clipped by the aperture.

As previously discussed under Degrouping Errors, a graded hole size is used to give increased light output in the central area of the screen where screen tolerances permit such increase. In the fabrication of the shadow-mask, the combination of graded hole size and graduated taper is obtained by the use of two different photographic masters, one located on each side of the mask. One master is graded, the second, having larger holes, is uniform. With the mask thus exposed, the etching process produces the desired grading and taper of the holes.

Bulb Design

The requirements of a bulb for a color tube, over and above those normally considered for the bulb of a black-and-white tube, are:

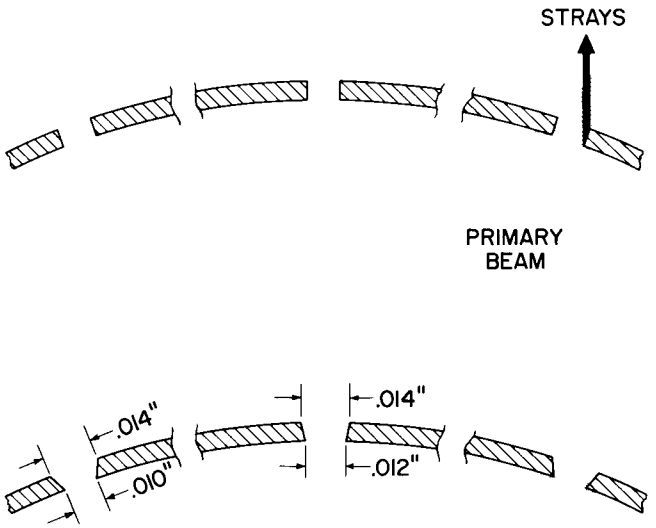


Figure 37. Advantages of Tapered Apertures in Reducing Stray Electrons

1. Means for mounting the mask in the panel section of the bulb.
2. Use of a separable bulb to facilitate screening.
3. More-accurate control of faceplate contour and neck alignment.
4. Use of a hard glass to minimize bulb distortion during thermal processing.

The first two items are essential structure requirements; the last two affect quality.

The mask is mounted by means of three tapered studs which engage the springs on the mask-frame assembly. In the metal tube, the 21AXP22, these studs were welded to the top wall of the panel section; in the glass 21CYP22 they are sealed to the wall of the glass panel. In both tubes, their positions are established by jigging from the faceplate contour and from the wall.

The two-piece bulb for the metal tube was made by flaring out the two metal sections of the bulb and welding them together at final closure by an inert-gas-shielded arc weld. The two sections were aligned by a dimple-and-groove system in the flanges of the two sections.

For the bulbs of glass tubes, it is customary to make the panel and funnel sections separately. Bulbs for black-and-white tubes are made by sealing these sections together, using an electric seal technique. This operation is usually done by the glass vendor. In the case of bulbs for color tubes, the two sections are sealed by a devitrifying frit after the screen is applied to the faceplate of the panel section. Because the glass itself is not melted in this sealing operation, the edges of the panel and funnel must be flatter and somewhat wider than those used for the black-and-white electric seal closure. The panel, being a pressed part, usually has satisfactory edge flatness; however, the method of fabricating the funnel is changed slightly. Rather than spinning the funnel and cracking off the glass at the top to form an edge, the glass is spun to a ring on the wide end of the funnel mold to form a relatively wide smooth edge.

The properties of the devitrifying frit are shown in Table I.

Table I

Properties of Devitrifying Frit
Used for Sealing the Funnel and Panel

Property	Condition of Frit	
	Vitreous	Devitrified
Softening-Point (deg C)	368	510 ± 10
Expansion (0-300 deg C) (10^{-7}/deg C)		96.5 ± 0.5
Log of Volume Resistivity (ohm-cm.)		
At 200 C		9.5
At 300 C		7.8

The important characteristic of this frit is that it is a soft-solder glass that devitrifies or changes into a ceramic phase upon sealing. The sealing process is not reversible, and once the seal is made, subsequent normal thermal processing will not soften it. The frit, which is made into a paste with amyl acetate and a nitrocellulose binder, is applied to the seal edge of the funnel by an extrusion machine. The seal is made by first allowing the frit to air dry, then positioning the panel on the funnel by means of an alignment jig, and finally, by sending the two parts through a sealing lehr. The thermal cycle involves raising the temperature of the parts to 440 C, holding this temperature for one hour, and then slowly cooling the parts. Upon completion of this process, the panel and funnel are sealed together and may be put through normal exhaust bakeout.

It is possible to combine the operations of sealing and exhaust. In this case, the gun assembly is first sealed in. Then a continuous thermal cycle is employed during which the frit is first sealed and then the exhaust pumping is started without the bulb cooling to room temperature. This technique results in appreciable saving in total processing time.

To design a tube that requires minimum correction and has optimum performance, the designer must consider the faceplate contour, alignment of the bulb parts, and reduction of bulb distortion.

The faceplate contour, measured with respect to the three studs, is important since it, together with the contour of the mask, establishes the "q" distance between the mask and faceplate required for optimum phosphor dot size and spacing. Tolerances of ± 0.023-inch at any point on the inner surface of the faceplate referenced from the three studs and panel sidewall has been maintained by the glass vendors on the 21CYP22.

Correct alignment of the panel and funnel is important because, for minimum purity correction, the electron beam positions in the deflecting yoke must coincide with the positions of the light sources in the screening light-

house. In the 21CYP22, accurate positioning is accomplished by relating the lighthouse referencing system to the system employed for aligning panel and funnel during envelope sealing. Fig. 38 illustrates the method used for referencing the panel on the screening lighthouse. The panel rests on its seal edge and is oriented by the V-pad and one of the plain molded-glass pads on the skirt of the panel. The same reference points are positioned, during panel-to-funnel sealing, against two ground pads near the seal edge of the funnel. These funnel pads are ground concentric with the inside of the neck by the glass vendor. Thus, when the gun is sealed into the neck, its position will correspond to the positions of the light sources in the screening lighthouse. The rotational orientation of the gun with respect to the panel is established on the gun-sealing machine with reference to the V-pad on the panel.

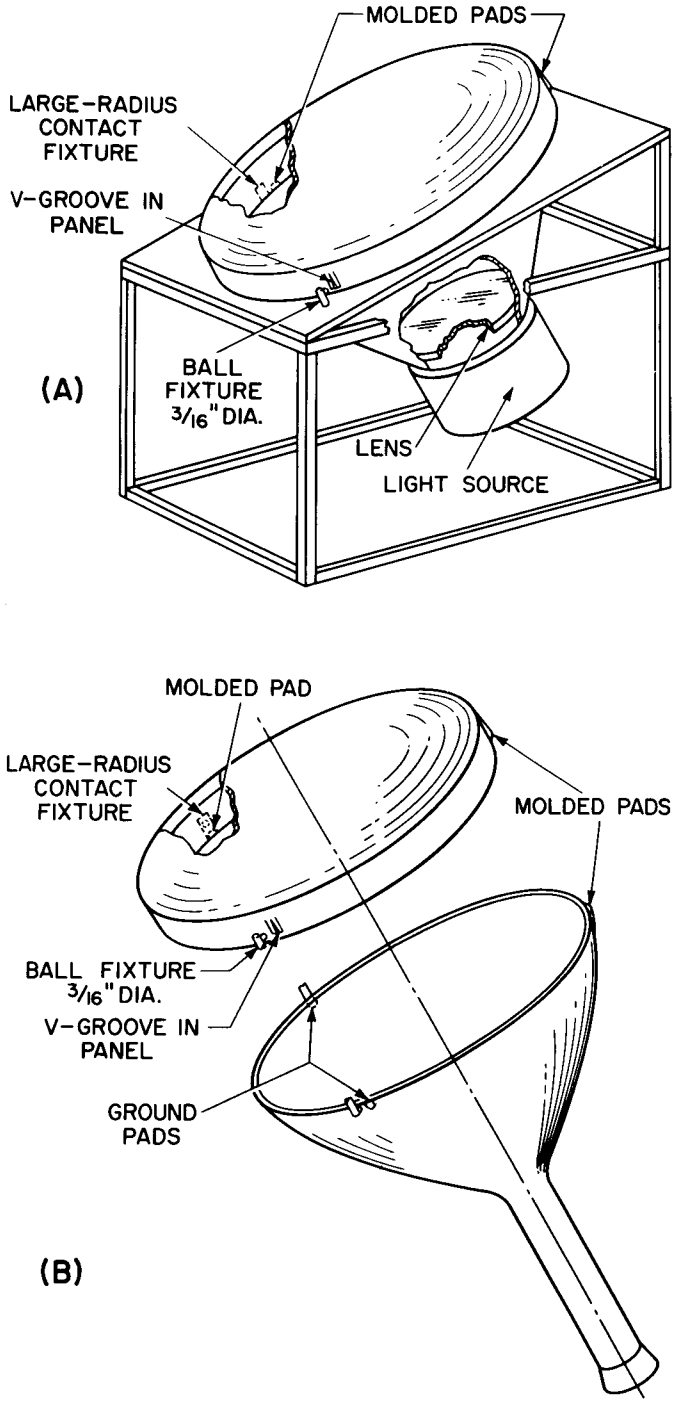


Figure 38. System Used for Referencing: (A) Panel on Screening Lighthouse; (B) Panel on Funnel

Distortions of the faceplate may be due to thermal effects and vacuum loading. These distortions normally reduce "q" or the spacing between the mask and faceplate. In the center of the screen this movement is relatively unimportant, but it becomes increasingly important toward the screen edge; the importance increases roughly as the tangent of the deflection angle.

With a metal envelope, the flatness of the welding flanges was the principal factor controlling faceplate distortion. If the flanges are not flat originally, they will be forced together by vacuum loading. This strain distorts the glass faceplate.

The glass tube, with its devitrifiable frit, eliminates the flatness problem; however, it is important that the strain point of the glass be high enough so that it is above the peak temperatures reached during thermal processing cycles.

Faceplate distortion is normally measured and referenced from the center of the faceplate. It has been customary to measure distortion at twelve equally spaced points near the edge of the faceplate. The average of the change in the twelve points is interpreted as average sag. This sag may be partially compensated for, if it is consistent, by the design of the radial lighthouse correction lens. The variations in values among the twelve readings, cannot be corrected within the tube. The differences from the average values and resultant misregister must be either allowed for as part of the screen tolerance or corrected by external magnets. Typical values of faceplate distortion are shown in Table II.

Table II

Typical Values of Faceplate Distortion

Bulb Type	Distortion	
	Average inch	Range inch
Metal Bulb	0.010	0.008
Glass Bulb, 432 C strain point	0.010	0.006
Glass Bulb, 455 C strain point	0.006	0.002

Gun Design

Features of the electrical design of the gun have been previously discussed.

The mechanical design of a triple gun requires that, in addition to the normal alignment of the parts within a gun, the three be properly aligned with respect to each other. The convergence of the three beams at the center of the screen requires that each gun be accurately tilted at an angle of about one degree to the tube axis. The three guns are individually beaded to two multiform glass beads 160 degrees apart. They are then placed on an assembly mandrel and welded together at the tops of the grids No. 4 and at the grids No. 3.

To give satisfactory convergence of the beams, the assembly mandrel should maintain the correct convergence angle within ± 10 minutes. The mandrels must also be designed so that they may be disassembled after the three guns are welded to each other.

Misalignment of the grids will cause misconvergence. The lens formed by grid No. 3 and grid No. 4 is especially critical in this regard. It has been found that an error of one mil in lens offset will produce 0.030 to 0.050 inch misconvergence. Although the pole pieces permit magnetic correction of convergence errors, excessive correction is to be avoided. Large corrections may cause beam distortions due to the strong correcting field. In addition, the triangular orientation of the three beams may be distorted because the orientation of the correcting fields may not be the same as the orientation of the convergence error.

The use of a three-bead assembly has certain advantages over the six-bead unit. Fig. 39 shows the three-bead assembly wherein the three guns are assembled as a unit and beaded together. This method simplifies the assembly operation and provides a more rigid mount than the six-bead unit with somewhat better alignment.

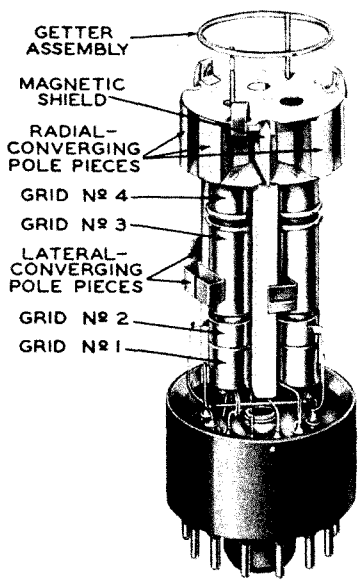


Figure 39. Three-Bead Gun Assembly

Stability of convergence during tube operation must be considered in gun design. Thermal expansion of the lower portion of the gun during the first 5 or 10 minutes of tube operation can cause beam movements of about 1/32-inch on the screen. This movement is usually quite repeatable within a given tube and is usually considered an insignificant quality defect. Thermal shifts should be considered in any major design modification.

Electrostatic charges on the inside of the neck opposite the open gun lenses may affect convergence. Because the charging of a bare glass neck is rather erratic, it has been found that an extension of the neck coating down over the lens formed by grids No. 3 and No. 4 of the 21CYP22 eliminates this charging problem. To minimize arcing from the end of this coating to grid No. 3, especially during spotknocking, a high-resistance iron-oxide coating extension is used.

COMPONENTS FOR SHADOW-MASK TUBES

Fig. 40 illustrates typical components for the 21CYP22. Because of the interaction of the tube and its components, certain features of their design must be carefully considered.

Yoke

The deflecting yoke for a shadow-mask color tube must deflect a bundle of three beams and, therefore, must handle a beam that is effectively one-half inch or so in diameter. The misconvergence at the edges of the screen is a function of deflection angle, the beam spacing, and the yoke design. The amount of misconvergence determines the dynamic correction needed, which, in turn, is the major factor in beam degrouping. (See reference on "Degrouping Errors.") It is, therefore, necessary to measure the yoke-tube degrouping pattern so that the proper corrections may be incorporated in the lighthouse correction lens.

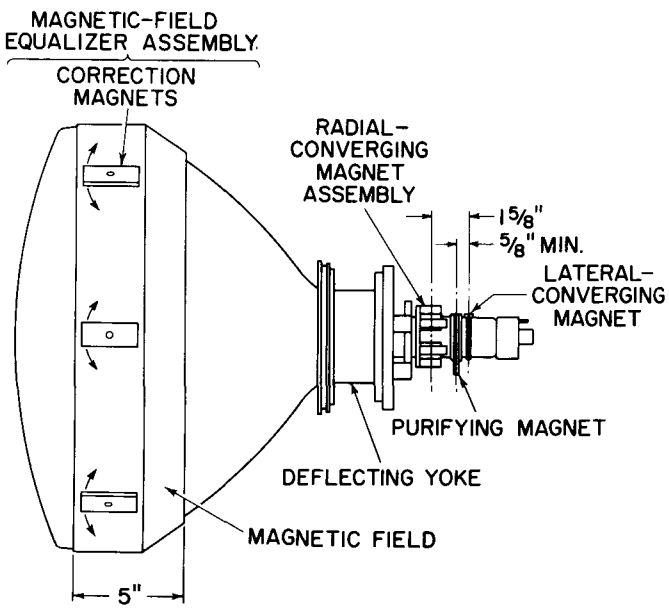


Figure 40. Relative Placement of Typical Components on 21CYP22

The electrical length of the yoke is important because it affects the radial correction to be used in the lighthouse lens. Because radial misregister is a function of both the yoke design and mechanical distortion of the bulb, the average radial misregister of a tube made with normal processing and with the bogey selected yoke should be made to establish the radial lens correction.

Coincidence of the horizontal and vertical centers of deflection of the yoke is essential for minimum misregister in a color tube. Although this coincidence is normally achieved automatically by the common core used for both the vertical and horizontal winding, the use of yoke shielding may upset this balance. This coincidence of the horizontal and vertical centers may be checked by observing the radial register on the screen as the yoke is rotated about the neck in a fixed axial location.

The back of the yoke is sometimes shielded to minimize the interaction of the yoke field with that of the convergence pole pieces. One type of interaction, which produces a red-green crossover, is illustrated in Fig. 41. The use of the "Y" shield in the pole piece assembly of the gun (as shown in Fig. 29) minimizes the yoke interaction and, hence, the need for yoke shielding.

Convergence Assembly

The design of the radial and lateral converging assembly should be such that it closely matches the internal poles (see Figs. 29 and 30). If the external and internal poles are not matched, the magnetic flux from the component will not be confined to the desired path and, as a result, interaction between beams and beam distortion may occur.

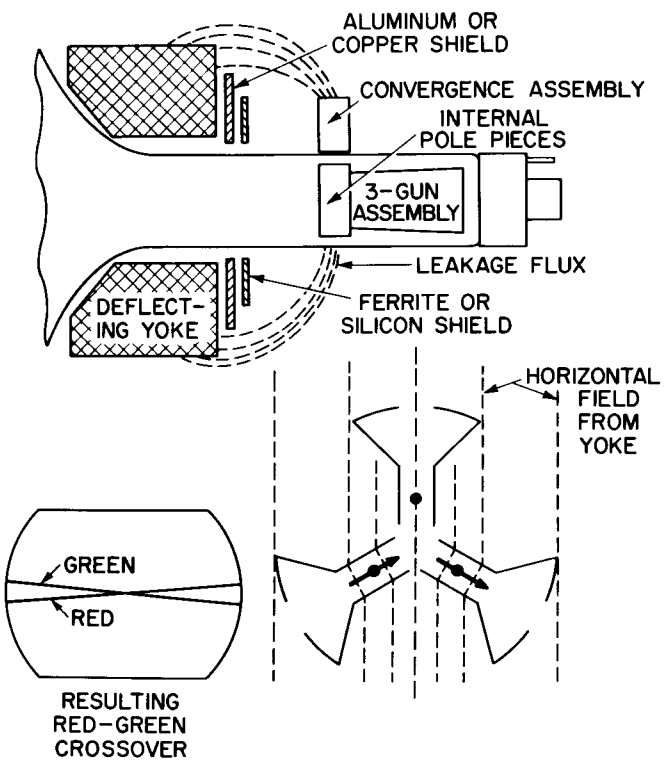


Figure 41. Red-Green Crossover Pattern Caused by Unshielded Pole Pieces

Purifying Magnet

The purifying magnet shifts the beams to compensate for mechanical or magnetic effects in a manner such that the beams are made to appear to originate from the color centers. Usually, the purifying magnet is located on the neck over the gun although "front end" or post-deflection purifying magnets may be used. The neck magnet is cheap and easily adjusted, but has the disadvantages of shifting the raster (thereby requiring centering in the deflection circuitry) and of distorting the beam if a high amount of correction is required. Front end magnets do not shift the raster or distort the beam, but they are difficult to adjust so that a uniform shift over the entire screen area is produced.

Equalizers and Shielding

Tube shielding minimizes the effects of magnetic fields. The equalizer magnets correct for misregister errors caused either by mechanical errors or by magnetic fields. The magnetic effects will be discussed later. The mechanical errors compensated for are those that would affect only localized areas of the screen. Bulb distortion, damaged mask, and incorrect dot size are some typical causes of localized misregister. The equalizer magnets mounted near the screen normally affect only a small area although the distribution of the magnetic field, and therefore its effect on register, may be modified by the distance from the tube and physical size of the magnet, and by the use of pole pieces. The pole pieces may be a part of the equalizer assembly or may be formed by the shield, metal cone, or internal frame and mask of the tube.

Improved control of all screening tolerances has made it possible to eliminate the equalizer magnets on recent tubes. The use of the shield and degaussing are sufficient to give good purity and uniformity.

OPERATIONAL CHARACTERISTICS

The requirements and the effects of tube operation are closely related to the tube design and, therefore, certain features in the application and usage of the tube must be understood by the tube designer. The special characteristics of shadow-mask color tubes include drive characteristics and magnetic effects.

Drive Characteristics

The production of a satisfactory black-and-white picture requires that the color tube maintain a constant color temperature over the entire range of the gray scale. Such a tube adjustment is also essential for a color reproduction to ensure that the color signal itself is the only factor that will cause the screen color to change. To achieve this condition, it is necessary that the ratios of the beam currents striking the red, green, and blue phosphor dots be constant. In the 21CYP22, the relative phosphor efficiencies are such that, to produce a white, a beam current division of approximately 50 per cent, 30 per cent, and 20 per cent is needed for the red, green, and blue phosphors, respectively.

In considering methods for obtaining this condition, there are two basic methods that may be used. The first, is to have a different video drive ratio applied to each gun. The video drive ratio would be such that the desired division of current is obtained. This method would produce a constant color temperature from low-lights to highlights. Such a system is shown in Fig. 42.

The second method would employ equal video drive for guns so that a common amplifier, for black-and-white, could be used. This system modifies the drive characteristic of each gun by changing its cutoff so that the desired ratio of beam currents is produced with the same input signal. Fig. 43 shows the drive characteristics of the 21CYP22 gun. By changing the cathode-to-grid-No. -2 voltage, which in turn affects the cutoff, the drive characteristic is altered to make the low-cutoff gun relatively easy to drive as compared to the high-cutoff gun. From Fig. 43, it may be seen that for any value of drive voltage the same ratios between the three gun currents are maintained. This system has two major disadvantages. The voltages needed for grid No. 2 and grid No. 1, or for cathode bias, may exceed those normally available in a television set. This condition exists if there is a wide range of cutoff voltages within the tube. The second shortcoming is the relatively low current obtained from the red gun at zero bias voltage. If the tube is driven to zero grid No. 1 voltage, the red gun will "give out" first and, as a result, a color shift will be noted in the highlights.

In present receivers, a combination of the two basic systems is employed. Fig. 44 shows the so-called "tapped video" system. On a bogey tube, the cutoffs are equal and the tapped output of the black-and-white amplifier provides the proportional cathode drive to give the desired current ratios. The color information is supplied to the grids No. 1.

The adjustment of the grid No. 2 voltages is made to compensate for variations in the cutoffs of the gun and in the current ratios needed to produce white in individual tubes.

The requirements of individual cathodes, grids No. 1, and grids No. 2 for this drive system place some limitation on gun design. Any consideration of a simplification of a mount structure that would use electrically common cathodes or grids must be balanced against the additional complexity of the video-drive circuitry.

Magnetic Effects

In a shadow-mask tube, register will be affected by

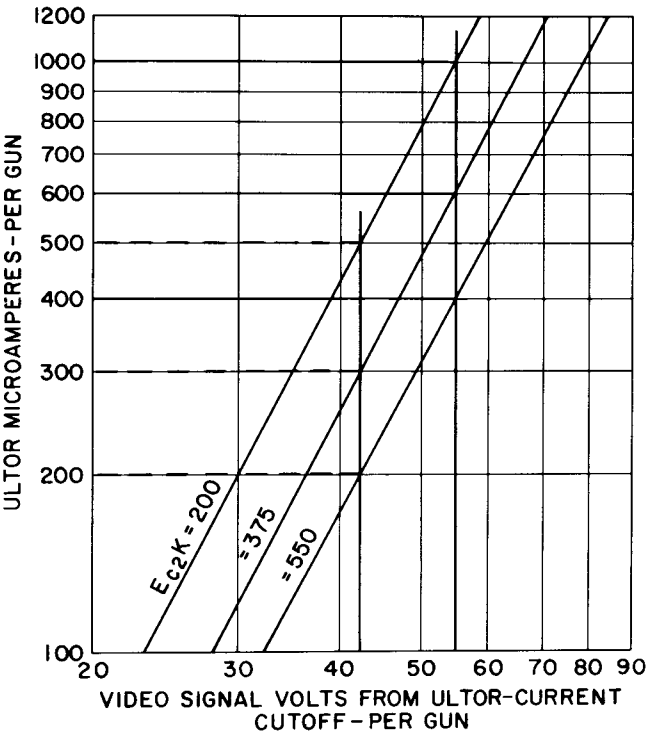


Figure 43. Adjustment of the Drive Characteristic by Changes in the Voltage Between Grid No. 2 and Cathode

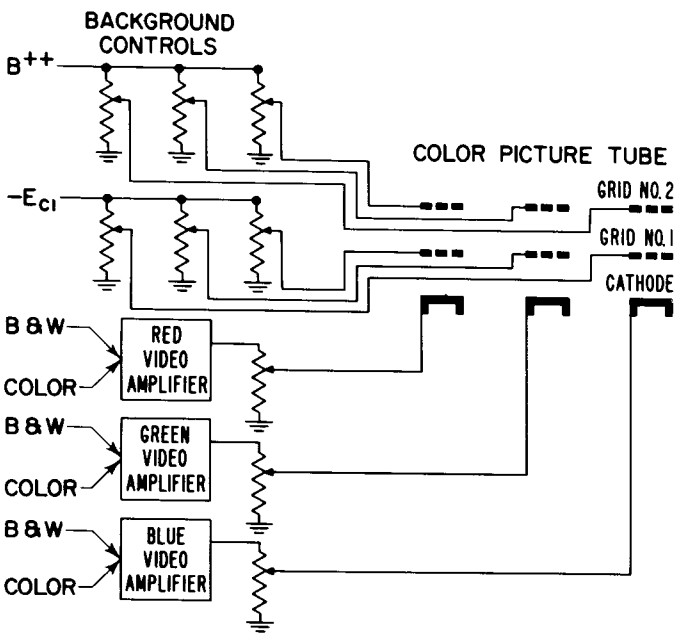


Figure 42. Individual-Video-Gain-Control System

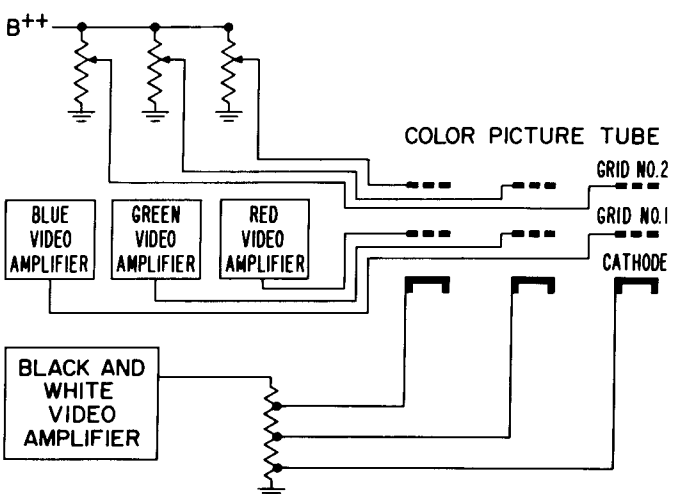
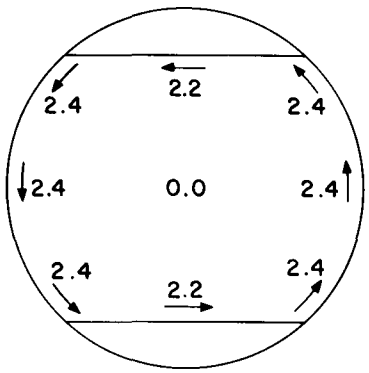


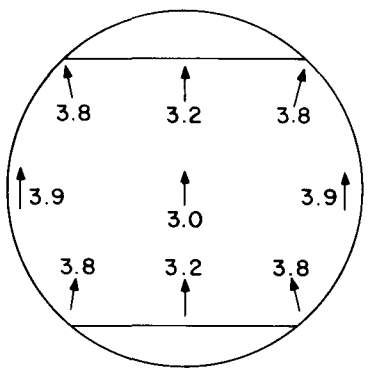
Figure 44. Tapped-Video System

a magnetic field normal to the path of the beam. The effect of a stray magnetic field on the beam prior to deflection by the yoke may be easily compensated for by the neck purity magnet. Stray fields acting on the scanned beam are more difficult to correct.

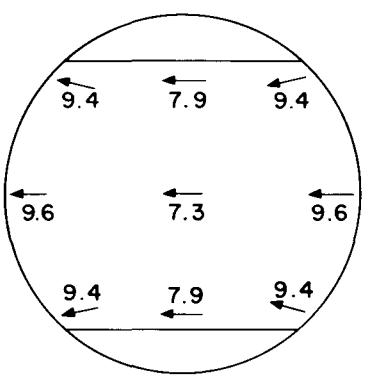
The effect of the earth's magnetic field on a theoretical color tube, similar to the 21CYP22, but possessing no magnetic parts to disturb the earth's field would produce a misregister pattern as shown in Fig. 45.



MISREGISTER (IN MILS) DUE TO AN AXIAL FIELD OF 0.22 GAUSS - TUBE FACING NORTH



MISREGISTER (IN MILS) DUE TO A HORIZONTAL TRANSVERSE FIELD OF 0.22 GAUSS - TUBE FACING WEST



MISREGISTER (IN MILS) DUE TO A VERTICAL FIELD OF 0.54 GAUSS

Figure 45. Misregister on a Nonmagnetic Tube With Geometry Similar to That of 21CYP22

The exact misregister would vary slightly in various geographic locations because of the different field strengths. The values shown in Fig. 45 are typical of the field in the Northern Hemisphere.

In practice, this misregister pattern is rather extensively modified by the magnetic parts of the tube or by magnetic materials used in the television receiver chassis or tube mounting. In particular, both the steel frame and steel mask have a great effect on the shielding and shaping of the earth's field. The misregister due to a typical value of the earth's field on a 21CYP22 is shown in Fig. 46.

The use of total shielding to eliminate these magnetic effects has been considered but in practice this method is not feasible. To be effective, such a shield would have to project well beyond the front of the tube to prevent the field from scooping into the tube. In a shield of any practical size, its effect in distorting and concentrating the earth's field at particular points might actually be more important than the reduction in the average field strength.

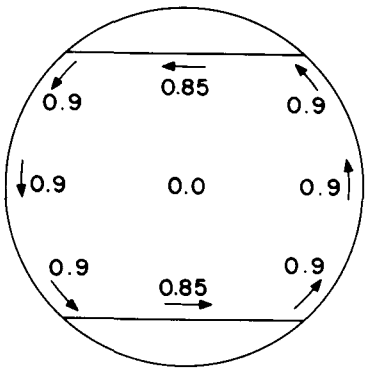
The use of high-permeability metals has been found unnecessary if proper degaussing procedures are followed. Degaussing a cold-rolled-steel shield by a 60-cycle magnetic field will give it magnetic shielding properties similar to those of a high-permeability material for a static field such as the earth's. Degaussing will also remove any residual magnetic effects that may be present in the tube or in surrounding metal parts.

The actual misregister pattern due to the earth's field of a given tube design and associated shield or mounting must be determined experimentally. Such tests are usually done in a magnetically shielded room or by means of large Helmholtz coils. With a shielded room in which the tube may be operated, the register pattern may be read on a given tube. By reading the register pattern again with the tube operating in the earth's field, and subtracting the two patterns, the misregister due to the field is obtained. The use of large Helmholtz coils to produce a uniform field of known intensity and direction over the volume of the tube may give more accurate readings because the field may be reversed to permit direct observation of the field effect and a higher field intensity may be used to magnify the effect.

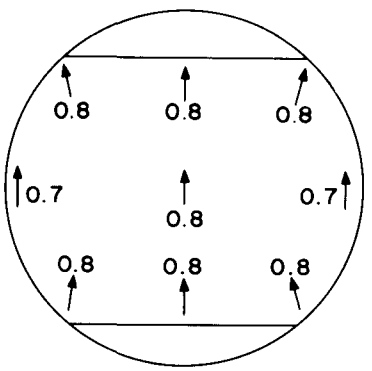
Compensation for the earth's field includes a number of considerations. First, the tube parts, especially those near the screen, i.e., the frame and mask, are made of magnetic material. Second, an external shield may be used to add to the shielding effect already provided by the tube parts. Third, the uniform portion of misregister caused by the vertical components of the earth's field may be removed by an offset in the lighthouse exposure. This offset is made such that the average misregister due to the earth's vertical field is zero. In the 21CYP22, the lighthouse offset is 0.040-inch. The remaining misregister must be allowed for in the screen tolerance or compensated for by means of the equalizer magnets.

DESIGN EQUATIONS

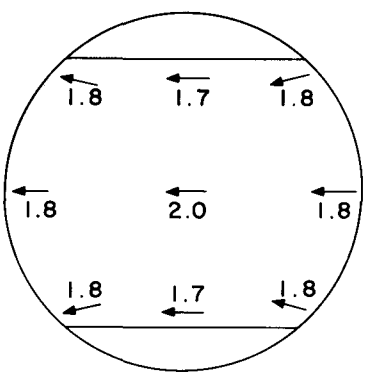
A summary of the more useful design equations that apply to shadow-mask tubes follows. With curved mask geometry, the particular value of some of the individual parameters is different at various parts of the screen. The equations apply to the center of the screen. They may apply to other portions of the screen if all other parameters in the equation are varied accordingly. A previous section describes the basic geometry and the general approach to the use of these equations.



MISREGISTER (IN MILS) DUE TO AN AXIAL FIELD OF 0.22 GAUSS - TUBE FACING NORTH



MISREGISTER (IN MILS) DUE TO A HORIZONTAL TRANSVERSE FIELD OF 0.22 GAUSS - TUBE FACING WEST



MISREGISTER (IN MILS) DUE TO A VERTICAL FIELD OF 0.54 GAUSS

Figure 46. Misregister on a 21CYP22 With Shield

Definitions (See Fig. 47)

- a = distance between aperture centers in the shadow mask.
- B = diameter of an aperture in the shadow mask.
- C = distance from the convergence plane to the deflection plane.
- D = minimum center-to-center distance between dots of the same color.
- d = distance from the center of a trio of dots to the center of a dot in the trio.
- g = distance from the center of the electron beam to the axis of the tube at the convergence plane.
- K = difference in diameters between tangent phosphor dots and the area of the dot excited by the electron beam.
- L = distance from the deflection plane to the phosphor screen.
- M = cross section diameter of the beam, at the deflection plane.
- q = distance from the shadow-mask to the phosphor screen.
- R = diameter of tangent phosphor dots.
- S = distance from the center of the electron beam to the axis of the tube, at the deflection plane.
- T = transmission of the mask.
- α = convergence angle of one of the beams in respect to the tube axis when converged at the center of the screen.
- λ = magnification or ratio of screen array size to the mask array size.

Equations

$$S = \frac{gL}{L + C}$$

$$q = (L - q)/\left(\frac{3S}{a} - 1\right) = \frac{La}{3S}$$

$$L = \frac{3\,Sq}{a}$$

$$D = R\sqrt{3}$$

$$d = \frac{R}{\sqrt{3}}$$

$$R = \frac{1}{\sqrt{3}}\left(\frac{3aS}{3S - a}\right)$$

$$B = \frac{a}{3}\left(\sqrt{3} - \frac{M}{s} - \frac{K\sqrt{3}}{R}\right)$$

$$M = S\left(\sqrt{3} - \frac{3B}{a} - \frac{K\sqrt{3}}{R}\right)$$

$$K = \frac{R}{\sqrt{3}}\left(\sqrt{3} - \frac{3B}{a} - \frac{M}{s}\right)$$

$$T = \frac{\pi}{18\sqrt{3}}\left(\sqrt{3} - \frac{M}{S} - \frac{K\sqrt{3}}{R}\right)^2 = 0.1008\left(\sqrt{3} - \frac{M}{S} - \frac{K\sqrt{3}}{R}\right)^2$$

$$T = \frac{\pi}{2\sqrt{3}}\left(\frac{B}{a}\right)^2 = 0.9069\left(\frac{B}{a}\right)^2$$

$$L = \left(\frac{S\sqrt{3}}{R} + 1\right)q$$

$$L = Rq\sqrt{3}/(R\sqrt{3} - a)$$

$$S = M/\left(\sqrt{3} - \frac{K\sqrt{3}}{R} - \frac{3B}{a}\right)$$

$$\alpha = \tan^{-1}\frac{S}{L}$$

$\lambda = \frac{L}{P}$ and for curved-mask geometry, $\lambda = \frac{R_f}{R_m}$ for a tube having constant magnification where R_m is the radius of the mask and R_f is the radius of the faceplate.[29]

21CYP22 Parameters

As an example, the various values of these parameters for the 21CYP22 are:

- S = 0.273 inch
- M = 0.050 inch for 500 microamperes; 0.090 inch for 2000 microamperes
- q = 0.535 inch at the center, and decreases towards the edge because of degrouping correction.
- L = 15.650 inches
- B = 0.012 inch at the center graded to 0.010 inch at the edge
- a = 0.028 inch
- R = 0.0168 inch
- K = 0.003 inch to 0.004 inch
- d = 0.0096 inch
- D = 0.029 inch
- T = 16.6 per cent at the center, 11.6 per cent at the edge
- g = 0.347 inch
- C = 4.125 inches
- α = 1° 0.5'
- λ = 1.034

COLOR PICTURE TUBE PROCESSING

The general outline of typical color tube processing is listed below. Only the major steps in processing and those that are different from black-and-white picture tube practice are described.

Panel Assembly

Mask. Etched in flat strip and formed to spherical shape.

Frame. Wrapped from strip and butt welded. Stabilized to minimize movement during tube processing.

Mask-Frame Assembly. Three springs are first welded to frame, then mask edge is welded to the frame. The frame is jigged from the spring and the mask is held by a contoured template. The assembly is then inspected for mask contour on an air gage.

Panel. Pressed by glass vendor. Inspected for glass quality and checked for contour on an air gage while jigged from the three studs.

Panel-Mask Assembly. Panel and mask-frame assembly put together and associated by the same identifying number to maintain this combination.

Screening Operation

The panel assembly is sent to the screen operation which includes the following major steps:

1. Panel is washed
2. A slurry of sensitized phosphor is spun on to the panel.

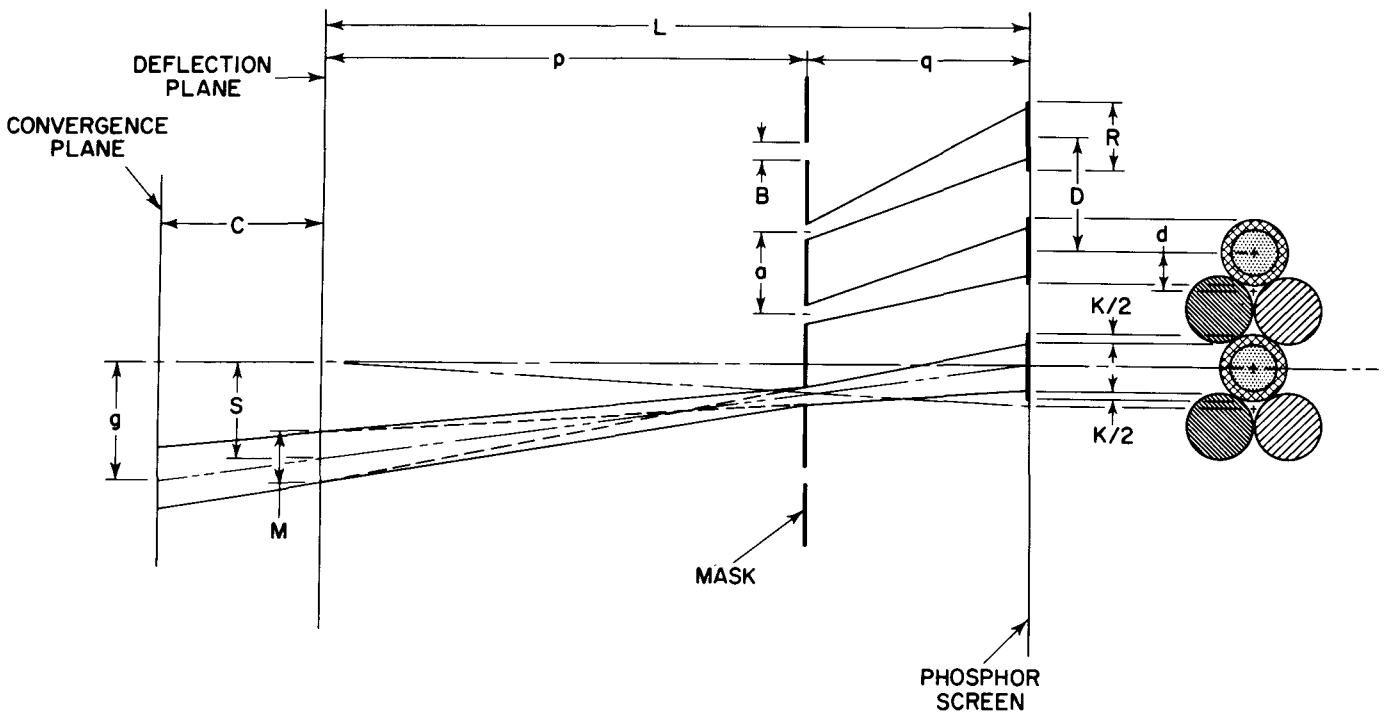


Figure 47. Tube Geometry

3. After screen drying, the mask is inserted into the panel and the assembly is placed on the lighthouse. The lighthouse[20] contains a high-pressure mercury-vapor lamp and a quartz collimator whose position with respect to the panel is the same as that of the electron beam of the same color gun in the deflection plane in the finished tube. The path of the light rays from the collimator is modified by the lighthouse correction lens to more nearly simulate the path of the electron beam in the completed tube. An exposure time of several minutes is used.
4. After exposure, the mask is removed and the screen is developed (washed) with water. Those portions that have not been exposed will be water soluble and will be removed leaving a dot array of the phosphor.
5. The process is repeated for the other two phosphors. Each time a screen is exposed, a lighthouse is used which has its light source in the position corresponding to the electron beam that will excite that array of dots.
6. The screen is filmed, aluminized, and baked.
7. The mask is reinserted to complete the panel assembly.

Funnel

1. Funnel inspection - check of neck runout with respect to the frit seal edge.
2. Washing and coating funnel with graphite.
3. Application of frit to the seal edge.

Frit Sealing

1. Prior to the sealing of the panel to the funnel, the electron shield and the contact springs which electrically connect the panel to the funnel coating are welded to the mask frame.
2. The frit seal is made by positioning the funnel and panel in the sealing jig and then heating the parts at 440 C for one hour.

Gun Seal

The sealing of the three-gun assembly is similar to that of a black-and-white tube except that the mount and the panel must have the correct rotational orientation.

Exhaust

The exhaust of a color tube and of a black-and-white tube are similar except that the heating and cooling rates for the color tube must take into account the frit-sealed bulb.

Basing, Aging, and Testing

These operations are similar to those for black-and-white tubes except that the color tubes require specialized tests associated with their color performance.

REFERENCES

1. Herold, E. W., "Methods Suitable for Television Color Kinescopes," Proc. IRE, Vol. 39, pp 1177-1185 (Listing and discussion of general categories), Oct., 1951
2. Fink, D. G., ed., Television Engineering Handbook, Sec. 5.14. (References to methods and specific tube types) McGraw-Hill Book Co., 1957
3. Barnett, G. F., et al., "A Beam Indexing Color Picture Tube - The Apple Tube," IRE Convention Record, pp 101-6, Part 3, 1956
4. Barnett, G. F., et al., "A Beam Indexing Color Picture Tube - The Apple Tube," Proc. IRE, V. 44, pp 1115-9, Sept., 1956
5. Payne, D., et al., "Recent Improvements in the Apple Beam-Indexing Color Tube," IRE Convention Record, pp 238-42, Part 3, 1957
6. Bond, D. S., et al., "Development and Operation of a Line-Screen Color Kinescope," Proc. IRE, V. 39, pp 1218-30, October, 1951
7. Code C1
8. Code C2
9. Code C3
10. Dressler, R., "The P. D. F. Chromatron--A Single or Multi-Gun Cathode-Ray Tube," Proc. IRE, V. 41, pp 851-8, July, 1953
11. Weimer, P. K., et al., "A 45-Degree Reflection-Type Color Kinescope," Proc. IRE, V. 39, pp 1201-11, October, 1951
12. Lafferty, J. M., "Beam Deflection Color Television Picture Tubes," Proc. IRE, V. 42, pp 1478-95, October, 1954
13. Law, H. B., "A Three-Gun Shadow-Mask Color Kinescope," Proc. IRE, Vol. 39, pp 1186-1194, Oct., 1951
14. Fyler, N., et al., "The CBS Colortron: A Color Picture Tube of Advanced Design," Proc. IRE, V. 42, pp 326-34, January, 1954
15. "Flat Tube for Colour TV," Wireless World, V. 62, pp 570-2, Dec., 1956
16. Gabor, D., et al., "A New Cathode-Ray Tube for Monochrome and Colour Television," Proc. IEE, Part B, V. 105, pp 581-606, Nov., 1958
17. Fyler, N., et al., "The Unipotential Mask-Focusing Colortron," IRE Convention Record, Part 3, pp 122-7, 1956
18. Lob, C. G., "General Electric Post-Acceleration Color Tube," IRE Convention Record, Part 3, pp 114-7, 1956
19. Ramberg, E. G., et al., "Focusing Grill Color Kinescopes," IRE Convention Record, Part 3, pp 128-34, 1956
20. Code C4

21. Goldstein, N. R., "Development of the Photographic Exposure Unit for Color-Picture-Tube Screens," Proc. of the 1958 Tube Techniques Conference, Session II, pp 81-8

22. Seelen, H. R., H. C. Moodey, D. D. VanOrmer, and A. M. Morrell, "Development of a 21-Inch Metal-Envelope Color Kinescope," RCA Review, Vol. 16, pp 122-139, March, 1955

23. Janes, R. B., L. B. Headrick, and J. Evans, "Recent Improvements in the 21AXP22 Color Kinescope," RCA Review, Vol. 17, pp 143-167, June, 1956

24. Smith, C. P., A. M. Morrell, and R. C. Demmy, "Design and Development of the 21CYP22 21-Inch Glass Color Picture Tube," RCA Review, Vol. 19, pp 334-348, Sept., 1958

25. Epstein, D. W., P. Kaus, and D. D. VanOrmer, "Improvement in Color Kinescopes through Optical Analogy," RCA Review, Vol. 16, pp 491-7, Dec., 1955

26. Code C5

27. Code C6

28. VanOrmer, D. D., and F. van Hekken, "Increased Light Output for Color Tubes through Optical Analogy," IRE Third Annual Electron Devices Meeting, Washington, D. C., Nov. 1, 1957

29. Code C7

30. Code C8

31. Code C9

32. Moodey, H. C., and D. D. VanOrmer, "Three-Beam Guns for Color Kinescopes," Proc. IRE, V. 39, pp 1236-40, Oct., 1951

33. Grimes, M. J., A. C. Grimm, and J. F. Wilhelm, "Improvements in the RCA Three-Beam Shadow-Mask Color Kinescope," Proc. IRE, Vol. 42, pp 315-26, Jan., 1954

34. Code C10

Circuit Troubles Caused by Unusual Tube Effects

W. E. Babcock

Harrison

Designers of electronic equipment usually consider the electron tube as simply another component in their circuits. Tube characteristics are defined by curves and tabulated data supplied by the manufacturer, and in general, tube performance can be calculated with reasonable accuracy from the published characteristics.

At some time or other, however, many circuit designers experience the feeling that electron tubes are part gremlin. This feeling is caused by some of the peculiar effects noted with certain tubes in certain circuits. These effects are not mentioned in the data supplied by tube manufacturers, nor are they described in any textbook on tubes or circuits.

Equipment designers have reported these peculiar effects to tube manufacturers at various times. In many cases, investigation of the effects by the tube manufacturers has determined their causes. In some cases, changes could be made in tube construction or design to eliminate the effects. In other cases, however, the effects have been found to be characteristic of the circuits used, and circuit changes are required to eliminate them.

This article discusses the causes of many unusual circuit problems in the hope that they may be eliminated in future circuit designs. It also describes some of the problems which face electron-tube manufacturers in making tubes at a reasonable cost to satisfy an almost infinite variety of applications. Continued cooperation between the equipment designer and the electron-tube manufacturer should alleviate these problems and should result in the production of equipment which has better performance and greater reliability.

SLEEPING SICKNESS

Perhaps the most attention on the part of both tube manufacturers and circuit designers in the last few years has been given to an effect termed "sleeping sickness." This effect has been most serious in computer circuits, where a tube may operate for long periods of time with its plate current cut off. In many cases, when a problem is put to the computer which requires the tube to pass a pulse of current, nothing happens. The tube just will not conduct.

The sleeping-sickness effect is caused by the formation of a resistance layer, known as the cathode interface layer, between the cathode base metal and the cathode coating. This resistance layer, which is very thin, acts as a dielectric for a capacitor in which the cathode base metal is one plate and the cathode coating is the other. This capacitor has a capacitance of the order of 0.01 microfarad and is shunted by the resistance of the interface layer.

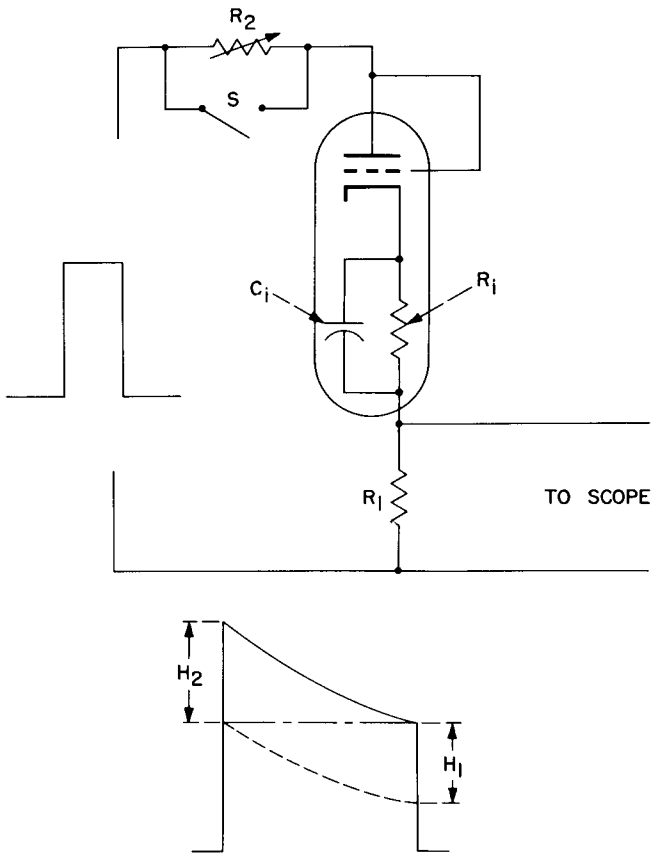


Figure 1. Effect of Cathode-Interface Layer on Plate Current

The effect of the cathode-interface layer is shown in Fig. 1. When a rectangular pulse is applied to the tube, the capacitance C_i acts as a direct short across the interface resistance R_i for an instant. The initial plate current of the tube, therefore, is determined solely by

the magnitude of the applied pulse and by the perveance of the tube. During the flat portion of the pulse, however, the "capacitor" charges up and the current begins to decay exponentially toward a value determined by the interface resistance.

Although the interface resistance is normally close to zero in a new tube, it can increase to several hundred or even a few thousand ohms after many hours of life with plate current cut off. The magnitude of the interface resistance can be determined by the use of a calibrated variable resistor R_2 in series with the pulse supply. The height of the trailing edge of the output pulse H_1 is first measured with R_2 set at zero. The value of R_2 is then increased until the leading edge of the output pulse H_2 has the same amplitude that the trailing edge had when R_2 was at zero. This value of R_2 is equal to the interface resistance.

Cathode-Interface Problem

Cathode interface is a direct result of the efforts of tube manufacturers to make tubes having a high value of emission at a reasonable cost. While the tubes are being evacuated during manufacture, the carbonate mixture which forms the cathode coating is broken down into oxides. Subsequently, during aging of the tubes, the oxides are broken down to form a surface layer of pure barium. A reducing agent is required at this step to assure that the oxides produce pure barium without liberating oxygen. This reducing agent consists of minute impurities (less than 1 per cent) added to the cathode base metal.

The impurity used as a reducing agent in most cathodes is silicon. Silicon forms the compound barium orthosilicate as an interface layer during periods when the barium is not being used to supply electrons for emission. In tubes designed especially for applications where "heater-only" or "plate-current-cut off" operation will be used, cathode material having an extremely low silicon content is used to limit the formation of an interface layer. These tubes can be operated for thousands of hours with negligible interface development. However, because the amount of reducing material available is small, factory processing of such tubes is much more difficult than that of tubes using cathodes which have a higher silicon content.

In applications requiring long periods of standby operation, either biased to cut off or with only heater voltage applied, it is advisable to use tubes designed for computer service wherever possible. If practical, it is recommended that some small amount of plate current be maintained at all times to delay the formation of interface.

Interface in Entertainment Tubes

In many cases it has been found desirable to use low-silicon cathode material in receiving tubes designed for home entertainment applications as well as in those designed for computer service. For example, low-silicon cathodes are used in the RCA-6SN7-GTB medium-mu triode, which is used as a horizontal-deflection oscillator in many television receivers. Most horizontal-oscillator circuits tend to exhibit frequency drift, and thus loss of "sync," if the oscillator tube develops interface.

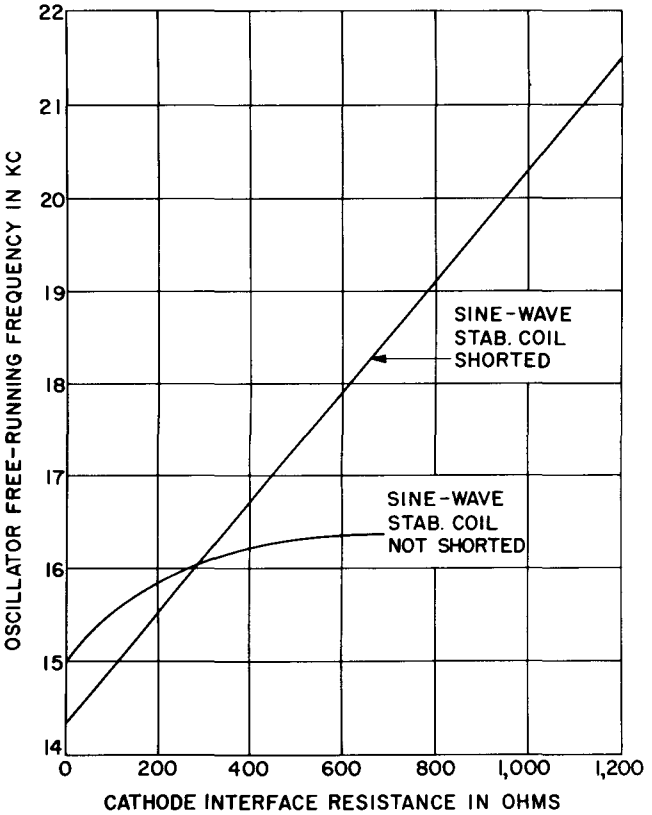


Figure 2. Free-Running Frequency of an Oscillator as a Function of Interface Resistance

Fig. 2 shows the free-running frequency as a function of the interface resistance for the synchro-guide type of oscillator circuit used in many television receivers. A similar curve can be obtained with the multivibrator type of horizontal oscillator. The upper curve, with the sine-wave stabilization coil shorted, shows the frequency drift when no attempt is made to compensate for tube variations. The second curve indicates the improvement in frequency stability achieved by circuit compensation. The compensation used in the synchro-guide circuit is shown in Fig. 3. The sine-wave stabilization coil, or ringing coil, tends to pull the oscillator frequency back to its proper value when changes in the tube or other components tend to change the frequency.

Although the use of sine-wave stabilization in the synchro-guide circuit reduced the frequency drift considerably, performance was still not considered satisfactory. Therefore, the cathode base metal of the 6SN7-GTB oscillator tube was modified to a low silicon content. Since this change was made, field troubles in this circuit due to cathode interface have been nonexistent, even after several thousand hours of operation.

BLACKOUT

Another effect, which in many respects is related to cathode interface has been called "Blackout." Another name used for this effect is "Whippany Effect." (So

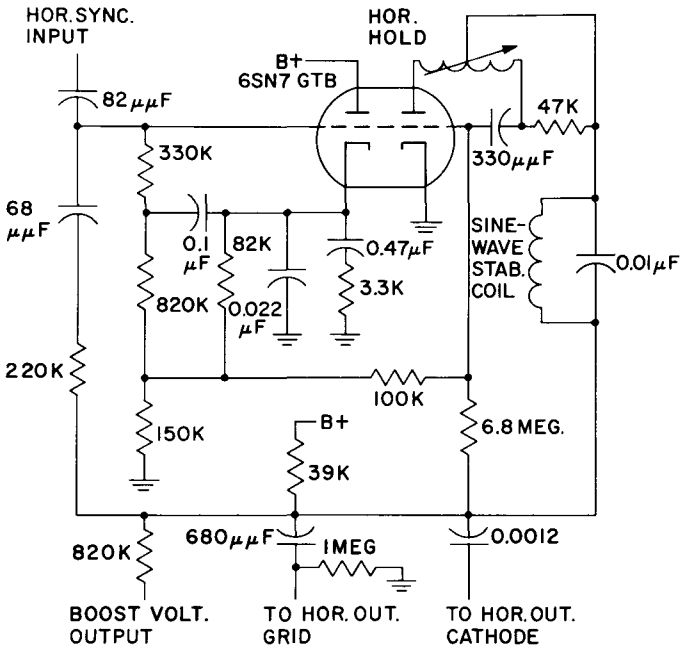


Figure 3. Compensation Used in Synchro-Guide Circuit

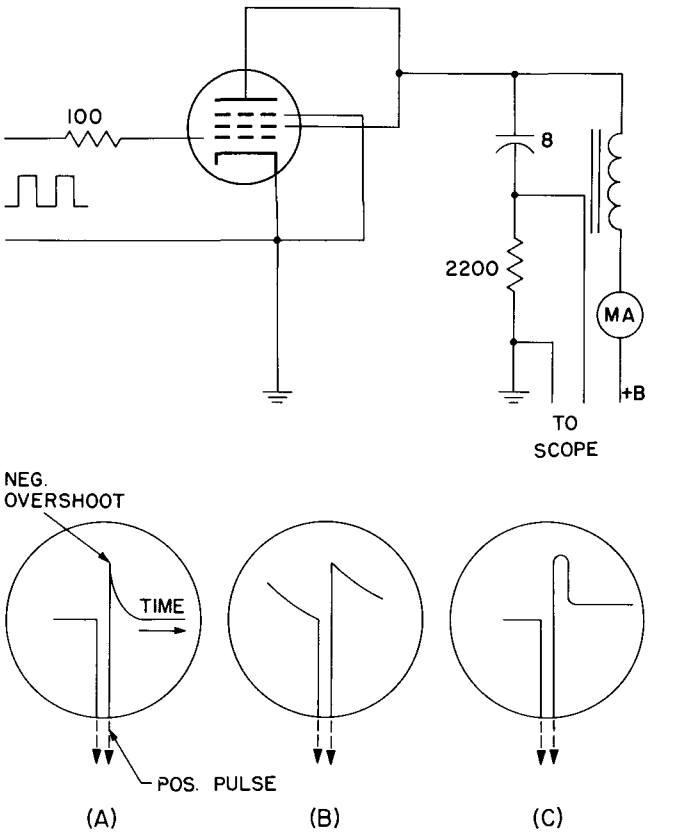


Figure 4. Blackout Test-Circuit and Oscilloscope Displays (A and B, poor tubes; C, good tube)

called because the effect was first noted by engineers of Bell Laboratories at Whippany, New Jersey.) This effect was first noted during the early days of World War II, when the transmitted pulse of a radar would get back into its receiver and cause it to go dead.

Blackout shows up only when the control grid is driven positive. A semi-insulating layer is sometimes deposited on the surface of the control-grid wires during tube manufacture. When the grid is driven positive, this plating acts as the dielectric of a capacitor in which one plate is the grid wire and the other is the layer of electrons collected on the plated surface. The tube then develops its own bias voltage internally as in the familiar grid-leak or grid-resistor method of biasing an oscillator or class C amplifier.

Figs. 4 and 5 show two methods for determining whether a tube has blackout. In the pulse method, shown in Fig. 4, the plate voltage of the tube under test is first adjusted for some given level of plate current, and a positive pulse is then applied to the control grid. If the tube is good, the plate current will probably rise slightly. However, if blackout is present, plate current will drop sharply. This change in plate current can be observed on an oscilloscope or a dc milliammeter. Oscilloscope displays obtained with a good tube and with two tubes having different degrees of blackout are also shown in Fig. 4.

In the dc method of testing for this effect, shown in Fig. 5, normal voltage is applied to the heater and a small positive bias voltage is applied to the control grid. If the tube under test is good, grid current will start to flow in about 10 to 15 seconds and will rise rapidly to its maximum value. If the tube displays blackout, however, the grid current will rise very slowly and may take as long as several minutes to reach maximum. The grid current increases slowly because the resistance layer on the control-grid surface has a negative temperature coefficient. The effect usually disappears completely in 5 to 10 minutes.

The negative temperature characteristics of the resis-sistance can produce some very annoying effects. For example, in the multivibrator type of horizontal-oscillator circuit, it may be impossible to hold the oscillator in sync for several minutes after the television receiver is turned on. Because the magnitude of the resistance varies as the tube heats up, the horizontal frequency also varies, and the picture drifts in the horizontal direction for several minutes every time the set is turned on. Similarly, frequency drift, or detuning, can occur in a local oscillator or in any class C amplifier stage because of this effect.

Tube manufacturers have not as yet developed a process for completely eliminating the development of this effect in tubes, nor have they been able to develop processing techniques which they can be sure will produce tubes completely free from blackout initially. Fortunately, however, normal processing techniques produce tubes which are generally free from this effect and most applications do not drive the control grid positive.

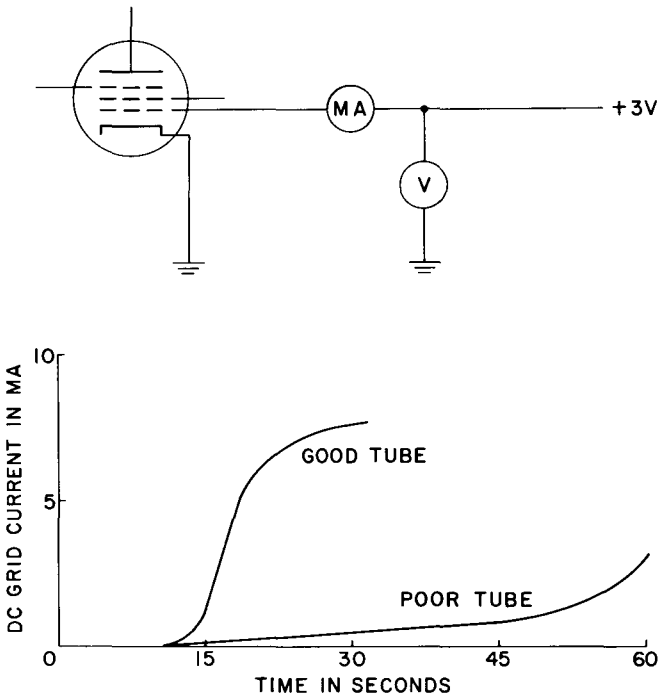


Figure 5. Direct-Current Test Circuit for Blackout and Resultant Curves

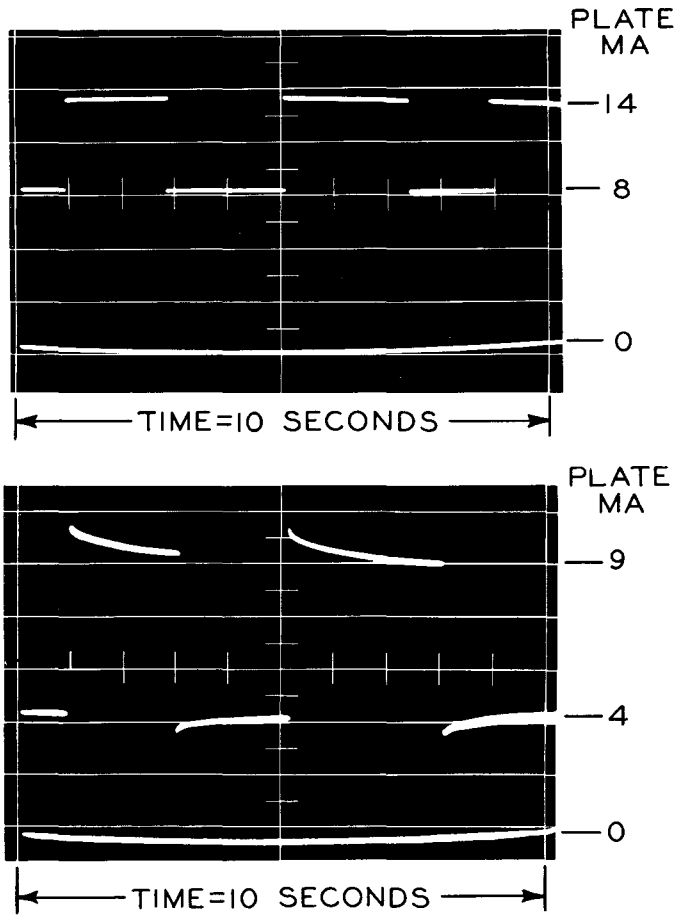


Figure 6. Output of Good Tube (top) and one With DC Shift (below)

DC SHIFT

Another effect, which appears to be similar to cathode interface and blackout, is known as "dc shift." In amplifiers using tubes having "dc shift," the amplifier gain is less for dc signals than for ac signals. In oscilloscope amplifiers, therefore, the usual ac calibration signal is not reliable if the oscilloscope is to be used for dc measurements.

The dc shift effect can be demonstrated dramatically by the use of a very-low-frequency square-wave signal, as shown in Fig. 6. If the tube is good, plate current remains at a constant level during the flat portion of the pulse. If the tube has dc shift, however, the plate current drifts about 5 to 10 per cent over a period of about 2 seconds.

The dc shift problem has been investigated by RCA in the tube laboratories at both Harrison and Princeton. It has been attributed by Dr. Nergaard of Princeton to the Sproull effect, which is caused by the formation of a donor depletion layer near the emitting surface of the cathode. In other words, dc shift is caused by the resistance of the cathode coating itself, which is normally called "bulk resistance." This resistance differs from cathode interface in that it has a time constant of the order of 1 or 2 seconds, while cathode interface normally has a time constant of the order of a few microseconds. This shift effect is inherent in tubes having close spacings, and can only be eliminated by the development of a new cathode material for receiving tubes.

Fig. 7A shows the equivalent circuit of a tube which displays dc shift in an amplifier circuit. If the average plate-current level is maintained constant in a particular application, and the departures from average are not too great, it may be possible to correct for most of the effect by a relatively simple compensating circuit such as that shown in Fig. 7B.

STRAY EMISSION

Stray emission in vacuum tubes can also cause some peculiar effects. The most familiar of these effects is probably the result of control-grid emission when an excessively large grid resistor is used — the tube "runs away" and usually destroys itself if no protective fuse is used. Stray emission from the screen grid (grid No. 2) in horizontal-deflection amplifier tubes is often responsible for loss of scan. In recent years, tube manufacturers have worked to reduce screen-grid emission in such tubes to satisfy field complaints of shrinking picture width, especially when the television receiver is operated at high line voltage.

The screen-grid current waveforms of two horizontal-deflection amplifier tubes are shown in Fig. 8. The emission current from the screen grid of tube B consumes power which would normally be used for deflection. As a result, insufficient power is produced for full deflection and the picture width is reduced. Simi-

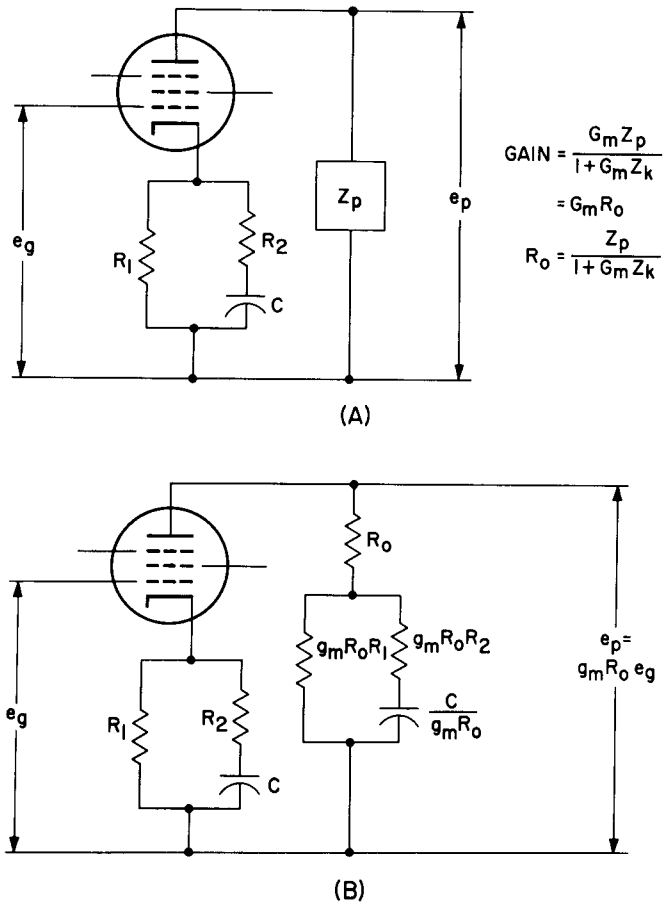


Figure 7. Equivalent Circuit of a Tube With DC Shift (A) and Compensating Circuit (B)[1]

larly, emission from the plate of the horizontal-deflection amplifer tube causes a power drain from the deflection system and a reduction in picture width. Although tube manufacturers have improved deflection tubes considerably and reduced the tendency toward screen-grid and plate emission, operation of tubes outside of ratings will seriously aggravate this problem.

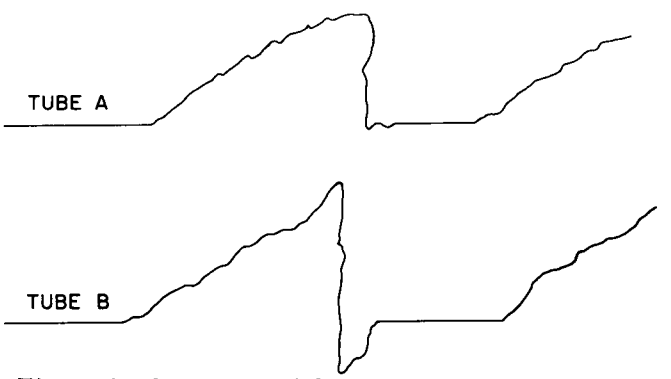


Figure 8. Screen-Grid Current Waveforms of Good Horizontal Deflection Amplifier Tube (A) and of One Having Screen-Grid Emission (B)

BULB OR MICA CHARGE

The deflection circuits of a television set also produce many peculiar effects. Some of the most mysterious effects are caused by bulb or mica charge, which produces jitter of portions of the picture. In horizontal-deflection tubes, such charge may produce streaks or jitter of the entire raster or, in the extreme case, an effect called "cogwheel" or "pie crust." Fig. 9A shows an unusually severe case of cogwheel; usually the effect is barely visible. In the vertical circuit, mica charge produces an effect known as "white line" or "bright line" as shown in Fig. 9B. It can also produce a "black line."

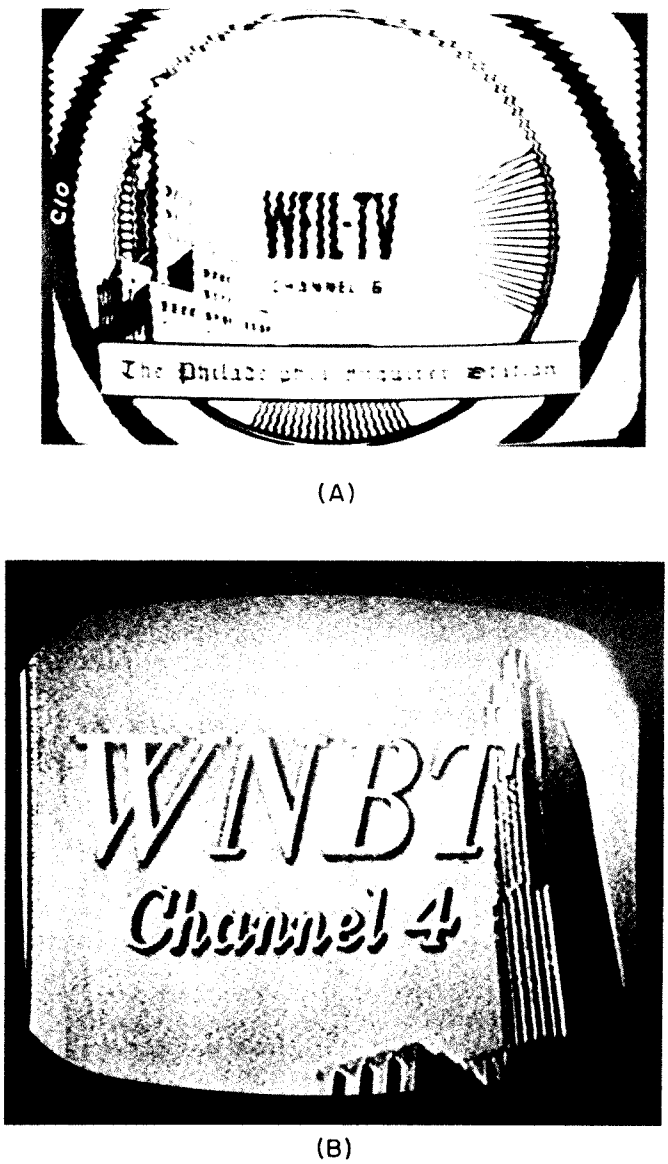


Figure 9. Severe Case of Cogwheel (A) and Bright Line (B)

Bulb charge can be minimized by the use of a carbonized bulb and by design of the tube structure so that electrons cannot strike the bulb. Carbonized bulbs are rarely used because loose carbon particles inside the tube can be more troublesome than the bulb-charge effect. Troubles due to mica charge are minimized by

mica designs incorporating slots which interrupt leakage paths and by the use of a high-resistance material such as alumina on mica surfaces. Equipment designers can help to reduce problems caused by bulb or mica charge by designing deflection circuits so that the peak plate voltages of the tubes are well within ratings.

SPOOK INTERFERENCE

"Spook" interference is another peculiar effect associated with the horizontal-deflection circuit. This type of interference, shown in Fig. 10, acquired its name because of its seemingly mysterious nature and because its cause eluded explanation for so long. The interference appears as a vertical line or band at the extreme left edge of the raster or, in many cases, may not be visible at all because of overscan of the raster or because it is in the blanked region. Sometimes, spook interference is picked up from a neighboring receiver and may move mysteriously back and forth across the screen; service technicians have referred to this behaviour as "windshield wiper" effect.

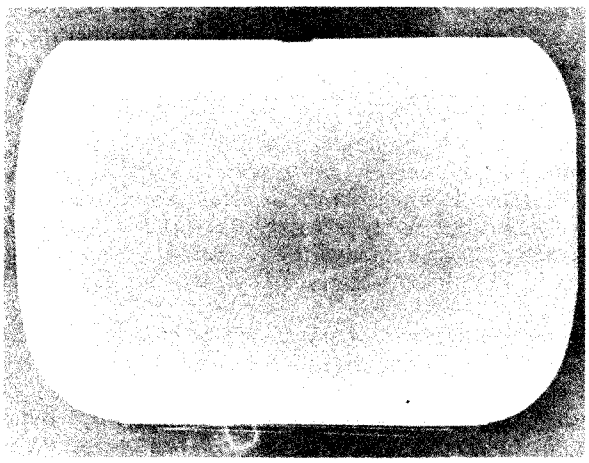


Figure 10. Spook Interference

Spook interference is generated by the damper tube. The plate current of this tube rises from zero to several hundred milliamperes in a very short time—about 0.1 microsecond. This rapidly increasing current produces many higher-order harmonics of the horizontal scanning frequency which lie within the television frequency bands. These harmonics often get into the sync circuits and cause picture instability.

It is impossible to eliminate the harmonics produced by the rise of current in the damper tube. However, if chokes are placed in the damper leads at the tube socket, the interference is limited to that radiated by the tube itself. It may also be necessary to put a shield around the damper tube to eliminate radiation from the tube structure.

SNIVETS INTERFERENCE

A very familiar kind of interference from the horizontal-deflection circuit is the "snivet" type, shown in Fig. 11. One possible cause of snivet interference is illustrated in Fig. 12, which shows the plate-current, plate-voltage characteristic of a deflection tube at zero bias voltage. When the plate current rises from zero to very high values, it follows a smooth curve. However, when it decreases from very high values toward zero, there is a discontinuity in the curve. This sudden change in plate current produces harmonics which can be picked up by the rf amplifier and produce interference.

Another theory holds that snivets interference is caused by a form of Barkhausen oscillation. This theory is logical because the plate voltage swings appreciably below the screen-grid (grid No. 2) voltage in many receivers. This condition is especially severe in modern flyback transformer designs which drive the plate voltage as far into the knee region as possible. An examination of the load line of a horizontal-deflection tube illustrates this phenomenon quite well. The most familiar load line to most engineers is that drawn for resistance-coupled amplifiers, which is simply a straight

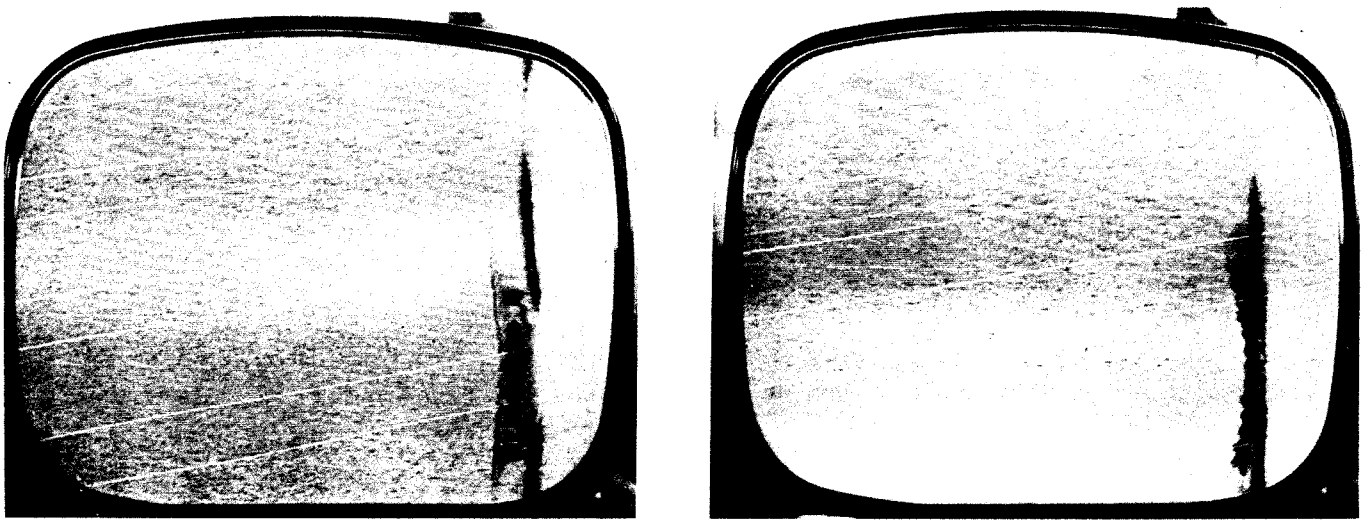


Figure 11. Typical Examples of Snivet Interference

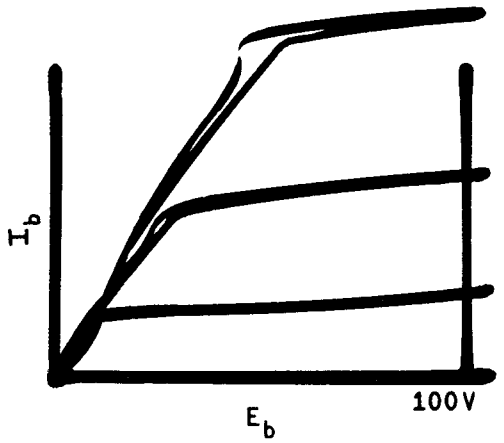


Figure 12. Horizontal-Deflection-Tube Plate Characteristic at Zero-Bias Voltage

line. If the load is reactive, the load line then becomes an ellipse. In contrast to these conventional load lines, Fig. 13 shows the load line of a typical deflection tube (RCA-25CD6-G) in a television receiver which exhibited very strong snivets. Is it any wonder that interference resulted?

REFERENCES

1. Code N1

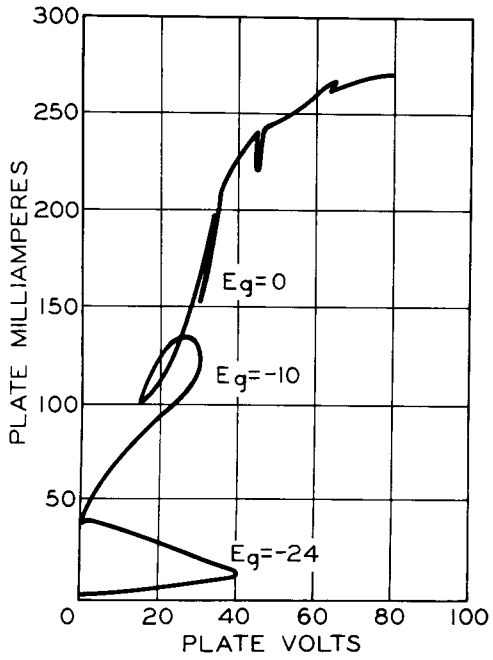


Figure 13. Load Line of a 25CD6-G Horizontal-Deflection Tube in a Receiver That Exhibits Strong Snivets

Some Factors Affecting the Design and Application of Small Power Tubes

J. W. Gaylord

Lancaster

INTRODUCTION

This chapter discusses the basic technical concepts used in designing electron tubes in which the electron transit time can be considered negligible. An arbitrary division between small- and large-power tubes can be made at a power output of 10 kilowatts. Although this discussion is primarily concerned with the design of tubes having power levels from a few watts to approximately 10 kilowatts, the same principles apply throughout the power spectrum.

Small-power tubes are used chiefly in communications equipment amplifying, generating harmonics, regulating voltage, modulation, switching, or generating signals by oscillation. The design of small power tubes entails the conversion of desired equipment-performance criteria into the mechanical and processing specifications of a physical tube.

Most equipment manufacturers, when specifying the requirements of a new tube, list the power output, bandwidth, and frequency objectives of their equipment design together with the limitations imposed by the available driving power, electrode voltages or currents, and environmental conditions. The tube-design engineer uses these equipment-performance objectives to arrive at a physical tube design by following a logical series of calculations and approximations. These steps are summarized in the following paragraphs and are developed in detail later in this article.

Design of Output Section

The output section of a power tube consists of the plate circuit and associated tube electrodes. The amplitude of the plate-voltage fluctuation and the fundamental component of plate current are related to functions of the frequency, bandwidth, and required power output of the equipment by two independent relationships which combine to determine the plate voltage and current. The output capacity of the lumped tube-circuit combination or the geometric shape of the distributed output circuit can then be found because they are functions of the plate voltage and operating frequency. The cathode area can also be determined from the required fundamental component of plate current by estimating the division of cathode current between tube electrodes and assuming a class of operation. This estimation of current division also permits the calculation of input power to the plate. The plate dissipation can be determined next by subtracting the power output of the tube, which is a function of the amplitude of the plate-voltage fluctuation and the fundamental component of the plate current, from the plate input power. At this point, the output section of the tube is fairly well determined if the physical design of the cathode (area requirement) and plate (dissipation requirement) are compatible with the output capacity or geometry requirements imposed by the plate voltage and operating frequency.

Design of Input Section

The input section of the tube is composed of the control-grid circuit and associated tube electrodes. The choice of operating class made during the design of the output section of the tube determines the bias-voltage level of operation. Relationships are available which permit the calculation of the geometric shape of the control electrodes as a function of the amplitude of input-voltage fluctuations and the bias-voltage level at the desired plate current can be determined from the amplitude of the input or driving voltage. If the driving power does not agree with the design objectives, alterations in the class of operation or control-electrode configuration, or both, can be made.

Tube Operation

The two independent relationships between the useful power output of the equipment, the frequency of operation, and the bandwidth which determine the plate voltage and current are quite simple, but require some derivation and explanation. The following section explains these relationships and presents the background material necessary for their interpretation.

Power tube performance can be calculated with reasonable accuracy by using the procedures outlined below. Two graphical methods of presenting the electrical characteristics of power tubes are presently used: (1) curves of plate current as a function of plate voltage at constant levels of control-electrode voltage, and (2) curves of control-grid voltage as a function of plate voltage at constant plate-current levels. With either type of characteristic presentation, the class of operation must first be chosen.

AMPLIFIER CLASSIFICATION

Classes of operation, which have been designated A,

AB, B, and C, are best described by comparing the relationship of the bias-voltage level E_{c_1} to the zero-current point of the transfer characteristic. A subclassification of 1 or 2 indicates whether control-grid current flows. The transfer characteristics for the various classes of operation are shown in Fig. 1. The angle Θ represents the portion of a 360-degree cycle during which current conduction across the tube structure occurs.

In general, the following rules apply. Large conduction-angle operation results in better linearity and power gain than small-angle operation. Small-conduction-angle operation results in increased plate efficiency and harmonic components in the plate-current pulse. [1]

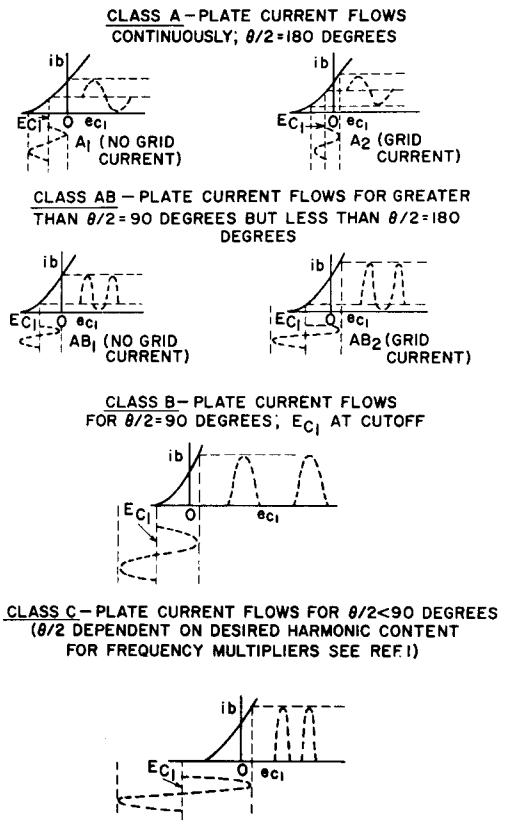

CLASS A – PLATE CURRENT FLOWS CONTINUOUSLY; θ/2 = 180 DEGREES
ib
Ec1
O
ec1
A1 (NO GRID CURRENT)
A2 (GRID CURRENT)
CLASS AB – PLATE CURRENT FLOWS FOR GREATER THAN θ/2 = 90 DEGREES BUT LESS THAN θ/2 = 180 DEGREES
AB1 (NO GRID CURRENT)
AB2 (GRID CURRENT)
CLASS B – PLATE CURRENT FLOWS FOR θ/2 = 90 DEGREES; Ec1 AT CUTOFF
CLASS C – PLATE CURRENT FLOWS FOR θ/2 < 90 DEGREES (θ/2 DEPENDENT ON DESIRED HARMONIC CONTENT FOR FREQUENCY MULTIPLIERS SEE REF. 1)

Figure 1. Transfer Characteristics for Classes of Amplifier Operation

Current Component Ratios

The relationship between peak, harmonic component, and average or dc plate current is required to calculate tube performance. The relationship between peak and average control-grid and screen-grid current is also required to calculate driving power and screen-grid dissipation. These relationships can be determined from an analysis of current wave shapes of the plate, screen, grid, and control grid. A graphical method or a Fourier analysis of the plate-current pulse can be applied.

Graphical Analysis of Constant-Current Curves

In the graphical method, the actual plate-current pulse shape is analyzed by a combination of graphical conduction-angle determination and analytical series expansion to arrive at the dc and fundamental components of plate, control-grid, and screen-grid current. Constant-current characteristics provide a means of graphically determining the instantaneous conduction angle. Because the shape of the plate-current pulse is not assumed, but is analyzed directly, the result is theoretically a more accurate determination of current components.

In the following example, a typical class C amplifier is operated at a dc plate voltage E_b of 900 volts and a dc screen-grid voltage E_{c_2} of 300 volts. The design procedure is as follows:

(1) Determine bias voltage: For class C amplifier operation, the half angle of plate-current conduction is less than 90 degrees. At slightly less than 90 degrees, the plate efficiency is not as great as at smaller angles; however, the power gain is higher. At an angle of 60 degrees, the second harmonic content of the plate-current pulse is at a maximum. Disregarding power gain, a compromise angle for best plate efficiency with least second-harmonic current is approximately 75 degrees. The required grid-No. 1 (bias) voltage E_{c_1} is determined as follows:

for a tetrode,

$$E_{c_1} = -\frac{\hat{e}_{c_1}\cos\ \theta/2 + \dfrac{E_{c_2}}{\mu_{c_2}}}{1-\cos\ \theta/2}$$

for a triode,

$$E_{c_1} = \frac{\left(-\hat{e}_{c_1} + \dfrac{E_p}{\mu_p}\right)\cos\ \theta/2 - \dfrac{E_b}{\mu_p}}{1-\cos\ \theta/2}$$

where

$\hat{e}_{c_1}$ = peak instantaneous control-grid voltage

E_{c_2} = screen-grid voltage

μ_{c_2} = screen-grid-to-control-grid mu factor

μ_p = plate-to-control-grid mu factor

$\theta/2$ = half angle of plate-current conduction

E_{c_1} = bias voltage

The peak plate current at which $\hat{e}_{c_1}$ occurs is approximately four times the dc plate current. The maximum dc plate-current rating of the 6816 is 180 milliamperes at maximum cathode-current loading; therefore, the peak plate current is approximately 720 milliamperes. On the constant-current curves for a screen-grid voltage of E_{c_2} of 300 volts, $\hat{i}_b$ occurs at a knee point in the curves where $\hat{e}_{c_1}$ is 16 volts and the minimum plate

voltage $\check{e}_b$ is 300 volts. The required bias voltage can then be expressed as follows:

$$E_{c_1} \approx -\frac{(16 \cos 75^\circ + \frac{300}{16})}{1 \cos 75^\circ} \approx -31 \text{ volts}$$

(2) Determine load line: The load line runs between the point where the peak control-grid voltage and peak plate current occur at the bend in the constant-plate-current line and the point where the bias voltage and plate-supply voltage coincide (see Fig. 2).

(3) Determine current components: A graphical-angle computer* is placed over the constant-current curves so that line OA passes through the e_{c1} point, line OG passes through the E_{c_1}-E_b point, and the guide lines are parallel to the load line. This transparent graph has lines OA, OB, OC, OD, OE, OF, and OG which are 15 electrical-conduction angle degrees apart. The values for i_b, i_{c_2}, and i_{c_1} are tabulated every 15 degrees as shown in Table I.

Table I

	i_b	i_{c_2}	i_{c_1}
A	700	80	90
B	680	50	80
C	600	10	50
D	400	0	10
E	150	0	0
F	15	0	0

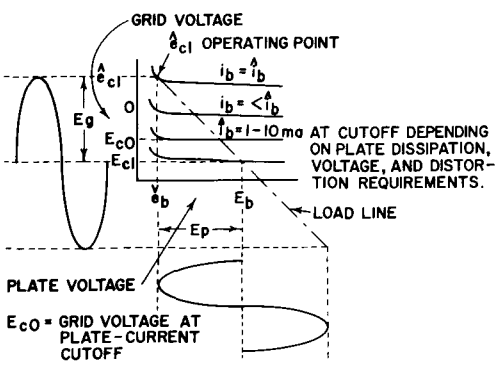

GRID VOLTAGE
OPERATING POINT
AT CUTOFF DEPENDING ON PLATE DISSIPATION, VOLTAGE, AND DISTORTION REQUIREMENTS.
LOAD LINE
PLATE VOLTAGE
Eco = GRID VOLTAGE AT PLATE-CURRENT CUTOFF

Figure 2. Constant-Current Curves for Typical Amplifier Load Condition

The actual half-conduction angle can also be read from the angle computer as 85 degrees by using one milliampere as the cutoff plate current, and progressing in angle from A to 2/3 of the 15-degree increment between F and G. The components of current can then be calculated from a series expansion of the current pulses by the following equations:

average or dc current, $I_b = 1/12\ (0.5A + B + C + D + E + F)$

peak of fundamental component, $I_p = 1/12\ (A + 1.93B + 1.73C + 1.41D + E + 0.52F)$

The results for the example cited are

I_b = 183 milliamperes
I_p = 314 milliamperes
I_{c_2} = 8.3 milliamperes
I_{c_1} = 15.4 milliamperes

(4) Determine power output, etc.: Power output and other tube performance characteristics are determined as follows:

Tube power output TPO = $1/2\ E_p/I_p$

$$= \frac{600 \times 0.314}{2} = 94 \text{ watts}$$

Useful power output = TPO x (output circuit efficiency) x (transit-time-efficiency*)

Plate power input PI = $E_b I_b$ watts = $900 \times 0.183 = 165$

Plate dissipation = PI - TPO = 71 watts

Tube Efficiency = $\frac{\text{TPO}}{\text{PI}} = \frac{94}{165} = 57$ per cent

Screen-Grid Dissipation = $E_{c_2} I_{c_2} = 300 \times 0.0083 = 2.5$ watts

Tube Driving Power = $E_g I_{c_1} = 46 \times 0.0154 = 0.71$ watt

Control-Grid Dissipation = $\hat{e}_{c_1} I_{c_1} = 16 \times 0.0154 = 0.25$ watt

Bias Dissipation = $E_{c_1} I_{c_1} = 30 \times 0.0154 = 0.46$ watt

Driver Power Output = $\frac{\text{Tube driving power}}{\text{Input circuit efficiency x Input transit time efficiency}}$

Fourier Analysis of Constant-Grid-Voltage Curves

A Fourier analysis of the current components can be used to determine the relationship between the average, peak, and harmonic components of current if it is assumed that the plate-current pulse is sinusoidal and the

*Application Bulletin No. 5, Tube Performance Computer, Eitel-McCullough Inc., San Bruno, California.

*The transit-time-efficiency factor has a value of approximately 1 up to 800 megacycles.

control-grid current pulse and the screen-grid current pulse have a squared sinusoidal form.[2] Table II summarizes the results of this analysis and indicates the normally chosen plate-conduction angles for various classes of operation.

Table III shows the grid-conduction angle, which is expressed as:

$$\theta_{g/2} = \cos^{-1}\left(\frac{-E_{c_1}}{-E_{c_1} + \hat{e}_{c_1}}\right)$$

where $\hat{e}_{c_1}$ is the peak control-grid voltage and E_{c_1} is the required bias voltage.

Choice of Operating Point

An operating point is chosen in the "knee" region of the plate-current characteristic at the current magnitude required by the equipment performance objective as shown in Fig. 3. The position of the operating point defines the tube performance as follows:

Tube power output = TPO = $1/2\ E_p I_p$

Plate power input = PI = $E_b\ I_b$

Plate dissipation = PI - TPO

Screen-grid dissipation = $E_{c_2}\ I_{c_2}$

Control-grid dissipation = $\hat{e}_{c_1}\ I_{c_1}$

Driving power = $(-E_{c_1} + \hat{e}_{c_1})\ I_{c_1}$ (for grid-drive circuit)

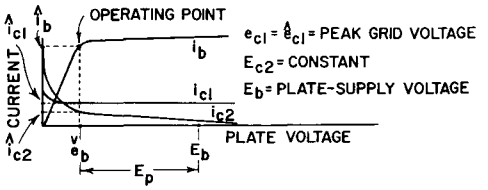


Figure 3. Constant Control-Grid Voltage Curves for Typical Amplifier Load Condition

Cathode Drive Considerations

If the amplifier is a cathode-driven device, the voltage reference is shifted. The only way to define operation properly is to indicate the circuit because, as shown in Fig. 4 and Table IV, either the grid or cathode can be at dc ground potential during cathode-drive operation.

I_{c_1} = dc control-grid current

I_b = dc plate current

I_p = fundamental ac component of plate current

E_{bb} = dc plate-supply voltage

E_b = dc plate-to-cathode voltage

E_p = amplitude of voltage fluctuation between plate and cathode

E_{c_1} = dc cathode-to-grid voltage

e_b = minimum amplitude of instantaneous plate-to-cathode voltage.

Table II

$\theta/2$ Half of total plate-conduction angle (degrees)	I_p amplitude			I_b dc plate current component	I_{c_2} dc screen-grid current component	Class of operation
	1st harmonic component	2nd harmonic component	3rd harmonic component			
90	$\hat{i}_b/2$	-	-	$\hat{i}_b/\pi$	$\hat{i}_{c_2}/4$	AB_1* or AB_2* B_1 or B_2
75	$\hat{i}_b/2.2$	-	-	$\hat{i}_b/3.7$	$\hat{i}_{c_2}/4.7$	C_1 or C_2
60	$\hat{i}_b/2.6$	$\hat{i}_b/3.6$	-	$\hat{i}_b/4.6$	$\hat{i}_{c_2}/5.8$	Doubler
40	$\hat{i}_b/3.6$	-	$\hat{i}_b/5.3$	$\hat{i}_b/6.9$	$\hat{i}_{c_2}/8.5$	Tripler

*Approximate solution suitable for most engineering problems.

θ = plate-current conduction angle

$\hat{i}$ = peak current

$\hat{i}_b$ = peak plate current

$\hat{i}_{c_1}$ = peak control-grid current

$\hat{i}_{c_2}$ = peak screen-grid current

I_p = amplitude of harmonic component of plate current

I_b = dc plate current

I_{c_1} = dc control-grid current

I_{c_2} = dc screen-grid current

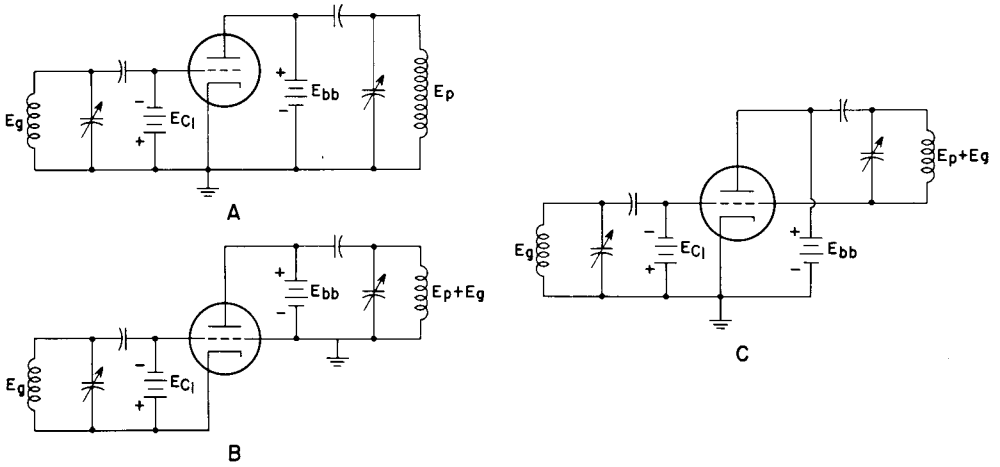


Figure 4. (A) Grid Drive Circuit, Cathode at DC Ground; (B) Cathode Drive Circuit, Grid at DC Ground; (C) Cathode Drive Circuit, Cathode at DC Ground

Plate power output = $1/2\ E_p\ I_p$ in all cases

Plate power input = $E_b\ I_b$ in all cases

Plate dissipation = $E_b\ I_b - 1/2\ E_p\ I_p$ in all cases

Table III

$\theta_{g/2}$ Half of total grid-conduction angle (degrees)	I_{c_1} dc control-grid-current component
90	$\hat{i}_{c_1}/4$
80	$\hat{i}_{c_1}/4.5$
70	$\hat{i}_{c_1}/5$
60	$\hat{i}_{c_1}/5.8$
50	$\hat{i}_{c_1}/6.9$
40	$\hat{i}_{c_1}/8.5$
30	$\hat{i}_{c_1}/11.2$
20	$\hat{i}_{c_1}/16$

This procedure provides a better understanding of the amplitude of the plate-voltage fluctuation E_p and the fundamental component of plate current I_p. The relationship of these parameters to the tube power output TPO is expressed as follows:

$$TPO = 1/2\ E_p I_p$$

This expression is the first of two independent relationships which determine plate voltage and current.

CIRCUIT OPERATION

The losses that occur in the plate and coupling circuits must also be considered so that the power delivered by these circuits to the load can be predicted. The power at the load is the equipment designer's requirement. Determining the circuit losses requires an understanding of the Q factor.

The Q Factor

The basic definition of Q is given by

$$Q = \frac{2\pi f\ (\text{stored resonant energy})}{\text{power dissipated}}$$

The Q of a lumped parallel circuit (Fig. 5A) is determined from this basic definition as follows:

$$Q = \frac{R_{shunt}}{X} \text{ at resonance}$$

where $X = \omega L$ or $1/\omega C$

The derivation of this equation is shown below because it deals with the application of the stored-energy concept used later to analyze distributed circuits.

The total energy stored in the resonant circuit is calculated by calculating the energy stored in either the capacitor or inductor during the period when no energy is stored in the other reactance.

For the capacitor,

$$dW = e\ dq = \frac{q}{C}\ dq \text{ because } e = \frac{q}{C}$$

$$W = \frac{1}{C}\int_0^{\alpha} q\ dq = \frac{d^2}{2C} = \frac{E^2C}{2}$$

where

W = energy

q = instantaneous charge

α = max. charge

Table IV

	Plate Power Input (Watts)	Plate Power Output (Watts)	Plate Dissipation (Watts)	Total Power Output to Circuit (Watts)	Driving Power Required (Watts)	Ep (Volts)
Case 1 Grid drive Cathode grounded	$E_{bb}\ I_b$	$1/2(E_{bb}-\check{e}_b)I_p$	$E_{bb}\ I_b$ $-1/2(E_{bb}-\check{e}_b)I_p$	$1/2(E_{bb}-\check{e}_b)I_p$	$E_g\ I_{c_1}$	$E_{bb}-\check{e}_b$
Case 2 Cathode drive Grid grounded	$(E_{bb}-E_{c_1})I_b$	$1/2(E_{bb}-E_{c_1}$ $\check{e}_b)I_p$	$(E_{bb}-E_{c_1})I_b$ $-1/2(E_{bb}-E_{c_1}-\check{e}_b)I_p$	$1/2(E_g+E_{bb}-E_{c_1}$ $-\check{e}_b)I_p$	$E_g(I_{c_1}+\frac{I_p}{2})$	$E_{bb}-E_{c_1}-\check{e}_b$
Case 3 Cathode drive Cathode grounded	$E_{bb}\ I_b$	$1/2(E_{bb}-\check{e}_b)I_p$	$E_{bb}\ I_b$ $-1/2(E_{bb}-\check{e}_b)I_p$	$1/2(E_g+E_{bb}-\check{e}_b)$ I_p	$E_g(I_{c_1}+\frac{I_p}{2})$	$E_{bb}-\check{e}_b$

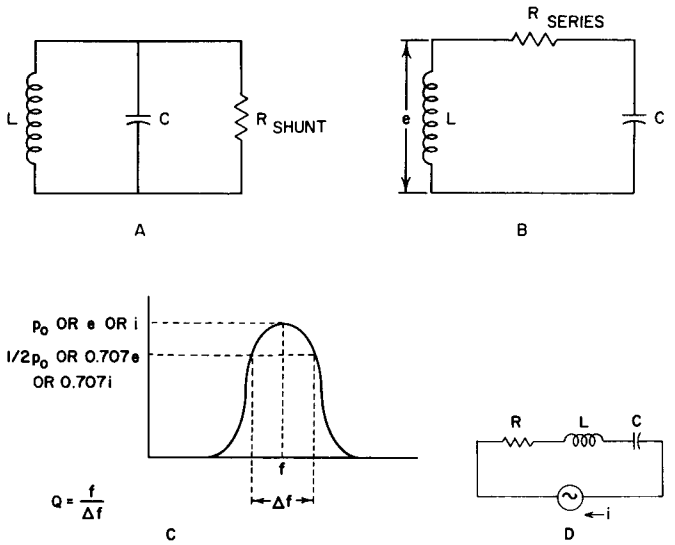


Figure 5. (A) Shunt Circuit; (B) Series Circuit; (C) Universal Resonance Curve; (D) Resonant Circuit to Relate Q and Bandwidth

When the capacitor is charged fully to α coulombs, the voltage across it is E sin ωt = E. At that time the rate of change of charge is zero and no current flows in L; therefore, no energy is stored in L. As a result, the total energy stored in the resonant circuit, which is all in the capacitor, can be expressed as $E^2C/2$.

The power dissipated per cycle per second can be expressed as

$$\frac{E^2}{R_{shunt}} \int_0^{1/f} \sin\ \omega t\ d\,\omega t = \frac{1}{2f}\ \frac{E^2}{R_{shunt}}$$

Therefore,

$$Q = \frac{2\pi f(\text{stored energy})}{\text{power dissipated}} = 2\pi \frac{1/2\ E^2C}{\frac{E^2}{2fR_{shunt}}} = 2\pi fCR_{shunt} = \frac{R_{shunt}}{X}$$

The Q of a lumped series circuit, as shown in Fig. 5B, is determined in a similar manner.

$$Q = \frac{X}{R_{series}}$$

at resonance where $X = \omega L$ or $\frac{1}{\omega C}$, the derivation is as follows:

for the inductor,

$$e = L\frac{di}{dt}$$

$$W = \int ei\,dt = \int L\,i\ \frac{di}{dt}\,dt = \int_0^I Li\,di = \frac{LI^2}{2}$$

When the current through the inductor is a maximum, its magnitude is I because $i = I\cos\omega t$. When the current is a maximum, the rate of change of charge is a maximum. This maximum value occurs when the charge on the plate of the capacitor is zero. Therefore, the total energy stored in the resonant circuit is all in the inductor and can be expressed as $LI^2/2$.

The power dissipated per cycle per second is

$$I^2\ R_{se} \int_0^{1/f} \cos^2 \omega t\,d\,\omega t = \frac{I^2\ R_{se}}{2f}$$

Therefore,

$$Q = 2\pi \frac{1/2\ LI^2}{\frac{I^2\ R_{se}}{2f}} = \frac{2\pi fL}{R_{se}} = \frac{X}{R_{se}}$$

As shown in Fig. 5C, Q is also related to the band-pass of the circuit under consideration as follows:

$$Q = \frac{f}{\Delta f}$$

where

f = frequency

Δf = bandwidth

The derivation of this relationship is included below because certain simplifying assumptions can be made which affect the calculated results for low values of Q.

In a circuit, such as shown in Fig. 5D, which has the basic condition for resonance, i. e.,

$$X_L = -X_C$$

$$j\omega_o L = j/\omega_o C$$

$$\therefore \omega_o = 1/\sqrt{LC}$$

where

$\omega = 2\pi f$ = angular frequency

$\omega_o = 2\pi f_o$ = resonant angular frequency

$\omega = \omega_o (i \pm \delta)$ at 1/2 power points

$i = i_o = \frac{e}{R}$ at resonance

$$i = \frac{e}{Z} = \frac{e}{R + j\left(\omega L - \frac{1}{\omega c}\right)}$$

$$i = \frac{e}{R} \frac{1}{1 + j \frac{\omega_o L}{R}\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)}$$

$$|i| = \frac{e}{R} \frac{1}{\sqrt{1 + \left(\frac{\omega_o L}{R}\right)^2 \left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)^2}}$$

At the 1/2 power points

$$i = i_o/2$$

Therefore,

$$i = \frac{e}{\sqrt{2}R} = \frac{e}{R} \frac{1}{\sqrt{1 + \left(\frac{\omega_o L}{R}\right)^2 \left(1 \pm \delta - \frac{1}{1 \pm \delta}\right)^2}}$$

$$1 + \left(\frac{\omega_o L}{R}\right)^2 \left(1 \pm \delta - \frac{1}{1 \pm \delta}\right)^2 = 2$$

$$1 \pm \delta - \frac{1}{1 \pm \delta} = \frac{R}{\omega_o L}$$

The result is the same for the positive or negative case. For the positive case,

$$1 + \delta - \frac{1}{1 + \delta} = \frac{R}{\omega_o L}$$

$$\delta^2 + \delta\left(2 - \frac{R}{\omega_o L}\right) - \frac{R}{\omega_o L} = 0$$

$$2\delta = \frac{R}{\omega_o L} - 2 \pm \sqrt{4 - \left(\frac{R}{\omega_o L}\right)^2}$$

Because

$$Q = \frac{R}{\omega_o L}$$

$$2\delta = \frac{1}{Q} - 2 \pm \sqrt{4 - \frac{1}{Q^2}} \approx \frac{1}{Q}$$

where the values for Q^2 are much greater than 1 and have a positive root.

Therefore,

$$Q \approx \frac{1}{2\delta} = \frac{f_o}{2\delta f_o} = \frac{f_o}{\Delta f}$$

The above information provides the basis for deriving the relationship which describes the circuit losses and predicting the useful power delivered by these circuits from the output gap in the tube to the load.

Circuit Efficiency

In the circuit shown in Fig. 6A, the tube is represented by a generator E. R_1 is the series generator impedance in the plate circuit, R_2 is the series load impedance in the load circuit, R_p and R_s are the series resistances of the coils, and X_p and X_s are the reactances. Therefore, $Q_p = X_p/R_p$ and $Q_s = X_s/R_s$.

The power lost P_{lost} in the transformer is expressed as follows:

$$P_{lost} = I_1^2 R_p + I_2^2 R_s = I_1^2 \frac{X_p}{Q_p} + I_2^2 \frac{X_s}{Q_s}$$

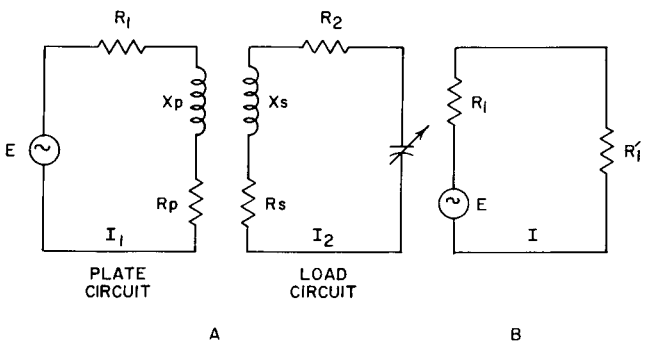


Figure 6. (A) Efficiency of Coupled Circuits; (B) Maximum Power Transfer Circuit

When a network of pure reactance, such as shown in Fig. 6B, is adjusted to the correct coupling position so that the tube is presented with the proper load impedance, the series impedance coupled into the plate circuit to obtain maximum power transfer to the load is R_1 because maximum power is transferred to the load when the generator impedance matches the effective load impedance. The proof is as follows:

$$\text{Power Output } P_{out} = I^2 R_1' = \frac{E^2 R_1'}{(R_1 + R_1')^2}$$

At maximum power output,

$$\frac{d(PO)}{d R_1'} = \frac{-2 E^2 R_1'}{(R_1 + R_1')^3} + \frac{E^2}{(R_1 + R_1')^2} = 0$$

$$R_1' = R_1$$

The plate circuit then appears as shown in Fig. 6B, where R_1 is equal to R_1'.

Therefore,

$$I_1 = \frac{E}{2R_1} \text{ under optimum loading conditions}$$

for maximum power transfer.

In addition, the voltage induced in the load circuit is ω MI.

Therefore,

$$I_2 = \frac{\omega M I_1}{R_2} = \frac{\omega ME}{2 R_1 R_2} = \frac{E}{2\sqrt{R_1 R_2}}$$

where M = mutual inductance

$$R_1 = \frac{(\omega M)^2}{R_2}$$

These values can be substituted for I_1 and I_2 in the equation for power lost.

$$P_{lost} = \frac{E^2 X_p}{4(R_1 + R_p)^2 Q_p} + \frac{E^2 X_s}{4(R_1 + R_p)(R_2 + R_s) Q_s}$$

If it is assumed that R_1 is much greater than R_p and R_2 is much greater than R_s, the power input P_{in} can be expressed as $P_{in} = I_1^2 R_1 = E^2/4R_1$

$$\text{The total efficiency } \eta = \frac{P_{out}}{P_{input}} = \frac{P_{input} - P_{lost}}{P_{input}}$$

$$= 1 - \frac{P_{lost}}{P_{input}} = 1 - \frac{X_p}{R_1 Q_p} - \frac{X_s}{R_2 Q_s}$$

Therefore,

$$\eta = 1 - \frac{Q_p'}{Q_p} - \frac{Q_s'}{Q_s}$$

where Q_p is the unloaded Q of the primary circuit (coil only), and Q_s is the unloaded Q of the secondary circuit (coil only). Q_p' is the loaded Q of the primary (tank circuit), and Q_s' is the loaded Q of the secondary (load circuit).

$$\text{If } \frac{Q_p'}{Q_p} \frac{Q_s'}{Q_s} << 1 \text{ since } (1-a)(1-b) = 1-a-b+ab$$

The total efficiency may be expressed as follows:

$$\eta = \left(1 - \frac{Q_p'}{Q_p}\right)\left(1 - \frac{Q_s'}{Q_s}\right) = \eta_{\text{plate circuit}} \eta_{\text{load circuit}}$$

Therefore, the tube power output TPO can be related to the useful power delivered to the load UPO as follows:

$$UPO = \eta TPO$$

The previously derived expressions can now be applied to determine the amplitude of plate-voltage fluctuation E_p and the fundamental component of plate current I_p from the equipment designer's requirements for power output, frequency, and bandwidth in the lumped-circuit case.

Determination of E_p and I_p for Lumped Circuits

Fig. 7 shows a lumped circuit and its equivalent circuit. From the relationships derived for Q,

$$R = \frac{E_p}{I_p} = 2\pi f L Ql = \frac{Ql}{2\pi f C} = \frac{1}{2\pi \Delta f C}$$

at loaded resonance

where

- Ql = loaded Q (the value of Q when the load R is receiving the desired power)
- R = plate-load resistance (transformed value of load resistance seen by tube output)
- E_p = amplitude of plate-voltage fluctuation
- I_p = fundamental component of plate current
- L = inductance across tube output
- C = capacitance across tube output (includes tube output, tuning, and stray capacitance in parallel)
- f = operating frequency
- Δf = bandwidth

Therefore,

$$E_p = \frac{I_p}{2\pi \Delta f C}$$

From the relationship derived for circuit efficiency:

$$I_p = \frac{2\ UPO}{E_p\left(1 - \frac{f}{\Delta f\ Q_u}\right)}$$

Substitution yields:

$$E_p = \sqrt{\frac{UPO}{\pi C\left(\Delta f - \frac{f}{Q_u}\right)}}$$

and

$$I_p = 2\pi \Delta f C E_p$$

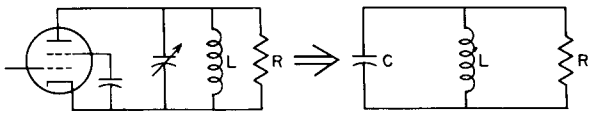


Figure 7. Equivalent Output Circuit

The value of C and unloaded Q (Qu) must be determined to complete the numerical calculation of E_p and I_p. C can be estimated from the physical size of the tube required to deliver I_p. The unloaded Q (Qu) is best estimated by experience rather than calculated. It can also be measured experimentally if a physical circuit is constructed. A value of 100 to 300 may be assumed for Qu in well-designed lumped circuits.

FREQUENCY EFFECTS

As the frequency of operation increases, it becomes difficult to resonate and couple power out of the output circuit simultaneously and efficiently because the inductance and capacitance must be reduced to maintain resonance as the frequency increases. The inductance cannot be reduced to a point which would prevent sufficient current coupling to the load. If voltage coupling is used, some increase in frequency can be achieved. Further increase in frequency would result in a reduction in output circuit capacitance. Eventually a frequency is reached where only the output capacitance of the tube remains. This arrangement prevents variable tuning and is therefore impractical. The minimum practical value of output capacitance of a nominal small-power tube circuit is approximately 10 micromicrofarads (tube-output, circuit-tuning, and stray capacitances).

The minimum practical value of circuit inductance is approximately 0.05 microhenries (1-turn, 1-inch-diameter coil).

Substituting these values in the resonance relationship yields:

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{0.05\times10^{-6}\times10\times10^{-12}}} = 225 \text{ megacycles}$$

Therefore, the maximum operating frequency of a nominal small-power tube using parallel-tuned lumped circuits is approximately 200 megacycles.

The effect of increasing frequency can also be interpreted in terms of bandwidth and circuit efficiency. The equation

$$\frac{E_p}{I_p} = \sqrt{\frac{L}{C}}\ \frac{f}{\Delta f}$$

is obtained by combining two previously derived expressions. According to the relationship $f = 1/2\pi\sqrt{LC}$, an increase in frequency requires a reduction in both L and C to maintain resonance.

Physical limitations eventually prohibit further reduction of either L or C. Further increases in frequency force a reduction of the other parameter and change the ratio of L:C. If this ratio increases because L is physically limited, the ratio $f/\Delta f = Ql$ must be reduced to hold E_p/I_p constant. A reduction of $f/\Delta f$ below 10 increases the bandwidth and results in the transmission of undesirable harmonics. If C is physically limited, the ratio L:C decreases causing the ratio $f/\Delta f = Ql$ to increase when E_p/I_p is held constant. This reduces the useful power output by decreasing the circuit efficiency:

$$\eta_{circuit} = 1 - \frac{Ql}{Qu}$$

The effect on tube structure of changing the ratio E_p/I_p is discussed under Perveance.

These limitations, which are inherent with lumped circuits, can be overcome by increasing the unloaded Q. The above expression for circuit efficiency indicates that if the value of Qu could be doubled, the same circuit efficiency could be obtained at twice the magnitude of Ql and, therefore, at twice the frequency. This increase in Qu is obtained by changing from a lumped to a distributed circuit.

Distributed Circuits

Typical values for Qu in coaxial resonators are shown in Table V.

Table V

Qu	Type of Circuit	Frequency
300	1/2 λ (conduction cooled)	500 megacycles
550	1/4 λ	400 megacycles
800	3/4 λ	1200 megacycles
1000	3/4 λ	2400 megacycles

Compared to an unloaded Q of 100 to 300 for lumped circuits, a considerable extension in frequency can be achieved by using distributed circuits.

The determination of E_p and I_p for the distributed circuit case requires the use of some simplified transmission-line theory and stored-energy concepts. These areas are briefly discussed in the following two sections.

Simplified Transmission-Line Theory

A transmission line is any device which tends to confine the electric and magnetic fields associated with a traveling wave. By confining these fields, the line gives direction to the traveling wave. The termination of a transmission line is highly reflective. The length of the line is determined so that reflected waves are in phase with incoming waves. The working equations for transmission lines are quite simple if no losses are assumed. Because the line parameters are a function of the direction of the wave, "sending" and "receiving" ends of the transmission line must be defined. When a series of

lines are resonant, the voltage and conjugate impedance at the junction of various lines must match. As in lumped circuits, a series of lines resonate when the reactance present in one direction is equal to the negative of the reactance present in the other direction (see Fig. 8).

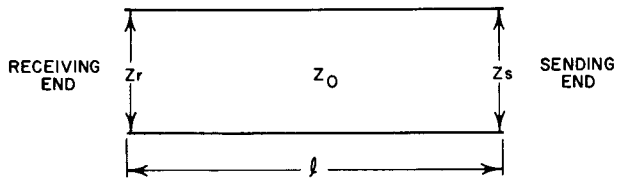


Figure 8. Basic Transmission Line Circuit

The basic transmission-line equations are as follows:

$$Z_S = Z_O \left[\frac{Z_r + j\, Z_O \tan \beta l}{Z_O + j\, Z_r \tan \beta l} \right] \qquad (1)$$

$$E_S = E_r \cos \beta l + j\, Z_O I_r \sin \beta l \qquad (2)$$

$$I_S = I_r \cos \beta l + j \frac{E_r}{Z_O} \sin \beta l \qquad (3)$$

where

$\beta = 2\pi/\lambda$
λ = wavelength
s = sending end
r = receiving end
Z = impedance
Z_O = characteristic impedance
l = length from chosen reference

Some interesting specific cases may be derived from these three basic equations. The characteristic impedance of a one-quarter-wavelength line at resonance may be derived from Eq. (2) or (3) as follows:

$$Z_O = -j \frac{E_S}{I_r}$$

$$Z_O = +j \frac{E_r}{I_S}$$

For an open-circuited receiver, from Eq. (2), (see Fig. 9):

$$E_S = E_r \cos \beta l$$

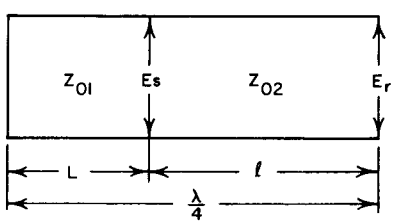


Figure 9. Open Circuit at Receiving End

At resonance, when Z_{O_1} equals Z_{O_2}, E_S equals $E_r \cos(\pi/2 - \beta L)$; therefore,

$$E_S = E_r \sin \beta L$$

For an open-circuited receiver, from Eq. (1),

$$Z_S = Z_O \left[\frac{Z_r + j\, Z_O \tan \beta l}{Z_O + j\, Z_r \tan \beta l} \right] = Z_O \left[\frac{1 + j \frac{Z_O}{Z_r} \tan \beta l}{\frac{Z_O}{Z_r} + j \tan \beta l} \right]$$

$$Z_S = -j\, Z_O \cot \beta l \text{ as } Z_r \longrightarrow \infty$$

For a short-circuited receiver, from Eq. (1),

$$Z_S = Z_O \left[\frac{Z_r + j\, Z_O \tan \beta l}{Z_O + j\, Z_r \tan \beta l} \right]$$

$$Z_S = +j\, Z_O \tan \beta l \text{ as } Z_r \longrightarrow 0$$

For cavities having various characteristic impedances, (at resonance, Z_1 equals Z_2), (see Fig. 10),

$$Z_1 = +j\, Z_{O_1} \tan \beta l_1$$

$$Z_2 = +j\, Z_{O_2} \tan \beta l_2$$

therefore,

$$Z_{O_1} \tan \beta l_1 = Z_{O_2} \tan \beta l_2$$

Stored Energy in Resonant Transmission Lines

In lumped circuits at resonance, the energy is considered to be stored in the electric field of the capacitor and the magnetic field around the inductor. This energy concept was used to calculate expressions for the Q of parallel and series lumped circuits. Analogously, the resonant transmission line stores energy in its associated electric and magnetic fields. This stored energy can be utilized in the definition of Q.

$$Q = \frac{2\pi f \text{ (stored resonant energy)}}{\text{power dissipated}}$$

to relate E_p and I_p to power output, bandwidth, and frequency.

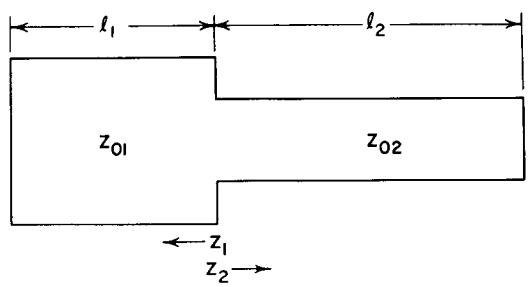


Figure 10. Resonant Circuit with Different Characteristic Impedances

Consider a resonant transmission one-quarter electrical wavelength long, as shown in Fig. 11. The energy

stored in a small element Δs of a line having capacitance C per unit length is defined as follows:

$$\Delta W = \frac{C\,\Delta s}{2}(E_p \cos \beta s)^2$$

where $\beta = 2\pi/\lambda$, E_p is the peak voltage across the line, and W is the stored resonant energy

$$dW = \frac{CE_p^2}{2}\cos^2 \beta s\, ds$$

$$\frac{d\beta s}{ds} = \beta$$

therefore,

$$\frac{d\beta s}{\rho} = ds$$

$$dW = \frac{CE_p^2}{2\beta}\cos^2 \beta s\, d\beta s$$

$$W = \frac{CE_p^2}{2\beta}\int_0^{\beta l} \cos^2 \beta s\, d\beta s$$

$$W = \frac{CE_p^2}{2\beta}\left[\frac{\beta s}{2} + \frac{\sin 2\beta s}{4}\right]_0^{\beta l}$$

$$W = \frac{CE_p^2}{4\beta}\left[\beta s + \sin \beta s \cos \beta s\right]_0^{\beta l}$$

$$W = \frac{CE_p^2}{4\beta}\left[\beta l + \sin \beta l \cos \beta l\right]$$

Characteristic impedance Z_o is defined as $Z_o = \sqrt{Z/Y}$.

The propagation constant is $\sqrt{ZY}$, which is equal to $\alpha + j\beta$ where α is the attenuation constant, and β is the phase constant. For low-loss lines,

$$Z_o = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}}$$

$$\sqrt{ZY} = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)}$$

therefore,

$$\beta = \omega\sqrt{LC}$$

$$\frac{\beta}{Z_o} = \frac{\omega\sqrt{LC}}{\sqrt{\frac{L}{C}}} = \omega C$$

$$C = \frac{\beta}{\omega Z_o}$$

Upon substitution for C, the expression for stored resonant energy reduces to

$$W = \frac{E_p^2}{4wZ_o}(\beta l + \sin \beta l \cos \beta l) = KE_p^2$$

The power dissipated per cycle is the power supplied by the tube divided by the frequency. The unloaded Q is then expressed as follows:

$$Qu = \frac{2\pi fW}{\text{power dissipated in circuit}}$$

The loaded Q becomes:

$$Ql = \frac{2\pi fW}{\text{power dissipated in circuit and load}} = \frac{2\pi fKE_p^2}{TPO}$$

Cavities having various Z_o sections and lumped reactances are treated in a similar manner. The stored energy in each section and lumped reactance is computed and added to obtain the total stored energy of the cavity system. See Douma,[3] Koros,[4] and Swift,[5] for a more complete discussion of the determination of stored energy in physical circuits.

Determination of E_p and I_p for Distributed Circuits

In this section, the information on transmission lines and Q is combined into an outline of the steps used to determine the amplitude of plate-voltage fluctuation and the fundamental component of plate current necessary to meet a specific useful-power output, bandwidth, and frequency requirement. In the steps that follow, Qu is the unloaded Q, Ql is the loaded Q, and K is a function of the physical properties of the resonant line.

(1) $Ql = \frac{f}{\Delta f}$

(2) Circuit efficiency = $\eta = 1 - \frac{Ql}{Qu}$

(3) Tube Power Output = TPO = $\frac{UPO}{\eta} = 1/2\, E_p I_p$

(4) Design the appropriate cavity for the frequency desired and calculate stored energy W in terms of E_p

$$W = f(E_p) = KE_p^2$$

(5) $Ql = \frac{2\pi f(\text{stored energy})}{\text{power dissipated}} = \frac{2\pi fKE_p^2}{TPO}$

(6) Rearranging the equation of step (5) and substituting yields:

$$E_p = \sqrt{\frac{UPO}{2\pi K(\Delta f - \frac{f}{Qu})}}$$

Substitution into Eq. (3) yields:

$$I_p = 4\pi K\, \Delta f E_p$$

It is important to note that for both lumped and distributed circuits the load impedance, E_p/I_p, presented to the tube must be decreased to get greater bandwidth.

At this point, it is of interest to compare the expressions for E_p and I_p derived for the lumped-circuit case to those parameters derived for the distributed-circuit case. This comparison results in the equivalence of

$$2K = C$$

Thus, K, which is a function of the physical properties of the tube and resonant line, is related to an equivalent lumped capacitance which is only a function of the physical properties of the tube and circuit combination.

PHYSICAL DESIGN CONSIDERATIONS

When the required amplitude of plate-voltage fluctuations and the desired harmonic component of plate current have been determined, the design engineer must design a physical tube structure which produces the required plate current and voltage (E_p and I_p) at reasonable driving power.

The following purely physical considerations must also be included in the design:

(1) compact size and low weight
(2) attractive appearance
(3) good life
(4) reproducible electrical and physical characteristics
(5) inherent low production cost
(6) a rugged mechanical structure
(7) stability of operation

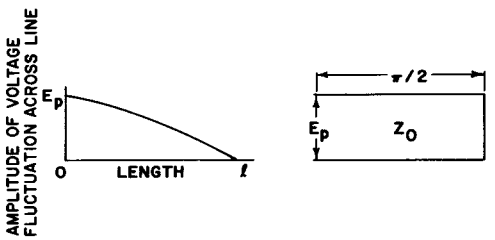


Figure 11. Voltage Along a Resonant Quarter-Wave Circuit

The remainder of this article deals with the conversion of the desired magnitude of E_p and I_p into the physical dimensions of the electrode structure. The problem of providing sufficient cooling for the electrodes is so specific for any given design that it is not practical to discuss it in general terms. From the practical standpoint, the temperature of the electrodes is a very real barrier to high-power performance. For this reason, the designer must consider the thermal properties of the tube design with as much attention as the electrical characteristics.

In the section which discussed power output and tube operation, graphical presentation of the electrical characteristics of the tube was used to determine the power output. These tube characteristics can best be converted to physical electrode dimensions by first considering the division of current which occurs between electrodes.

Current Division

The problem of determining the cathode current that flows to the plate of a diode and the plate and control grid of a triode is relatively simple compared to determining cathode current in tetrodes and pentodes. The pentode structure is rarely used in power-tube design because of added cost, circuit complexity at high frequencies, and increased output capacitance. Therefore, the following discussion covers the current division in the tetrode structure only.

A typical tetrode characteristic is shown in Fig. 12. The cathode-current equations discussed later provide the indicated cathode current. The sum of the plate, screen-grid, and control-grid currents at any plate potential must equal this cathode current (primary and secondary emission from the control grid, screen grid, or plate are neglected). Because no theoretical method of predicting the division of cathode current has yielded reliable results, only empirical means are discussed below.

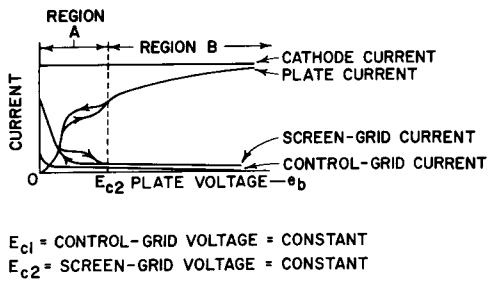


Figure 12. Typical Tetrode Current Characteristics

Current Division in Region A - Special precautions must be taken to ensure current stability in region A to permit tube operation where a portion of the plate voltage swing is below screen-grid potential. Regeneration resulting from a negative slope in the plate current can cause extraneous oscillations. [6]

Current Division in Region B - All tetrode structures exhibit stable current characteristics in this region.

Empirical Data on Cathode-Current Division

Figs. 13 thru 17 show the ratio of plate, screen-grid, and control-grid current to cathode current at E_b = 140 I_b, $E_b = 1/2\ E_{c2}$, $E_b = E_{c2}$, $E_b = 3/2\ E_{c2}$, and $E_b = 3\ E_{c2}$ for typical beam power tetrodes, parallel-plane tetrodes, and cylindrical tetrodes over a range of control-grid voltage from -100 to 500 volts. Data were taken at one or more screen-grid potentials from RCA types 7213, 7214, 1649, 6816, 6146, and 7094.

The division of current is strongly affected by the presence of space charge in the electron beam of the tube. [7]

A properly designed tetrode structure, which focuses the electrons emitted by the cathode into beams, (1) prevents primary cathode current from being intercepted by the screen grid, and (2) forms a space charge which prevents secondary electrons (emitted by the plate) from reaching the screen grid.

This reduction in screen-grid currents results in:

(1) Low screen-grid dissipation, which increases the total efficiency of the tube.

(2) Increased plate current at low plate potentials, which provides more efficient tube operation because of the added available plate-voltage swing.

(3) Stabilized screen-grid currents, which allow the screen-grid potential to be obtained from series screen-grid dropping resistors.

(4) High screen-grid voltage ratings, which result in increased power output.

(5) Low screen-grid temperature, which results in less primary screen-grid emission.

(6) Reduction in screen-grid secondary emission and the resulting operational instability associated with negative screen-grid current.

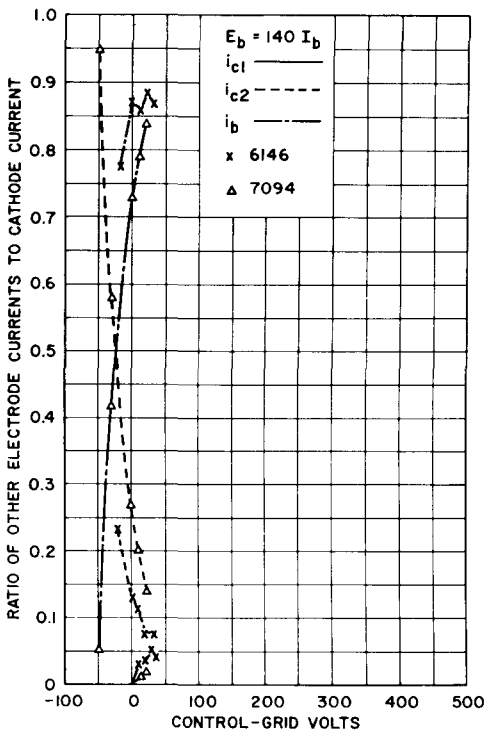


Figure 13. Current Division in Typical Tetrodes

The reduction of primary electron interception by the screen grid is accomplished by adjusting the equal potentials in the control-grid region so that during the conduction part of the grid cycle, the focus shifts from the region between the control grid and the cathode to some point near the screen grid. During this shift, the beam width at the screen grid must be less than the distance between the screen-grid lateral wires. This focusing action can be obtained by proper adjustment of the electrode structure while still preserving the magnitude of plate current. The pitch between grid wires, the degree of grid-wire alignment, the spacing between the active areas of the electrodes, and the control-grid driving conditions are adjusted to produce the desired plate current and focusing action simultaneously. Otto Schade[6] describes the lens action of the grids and a test to determine the relative effect of geometrical changes on screen-grid current.

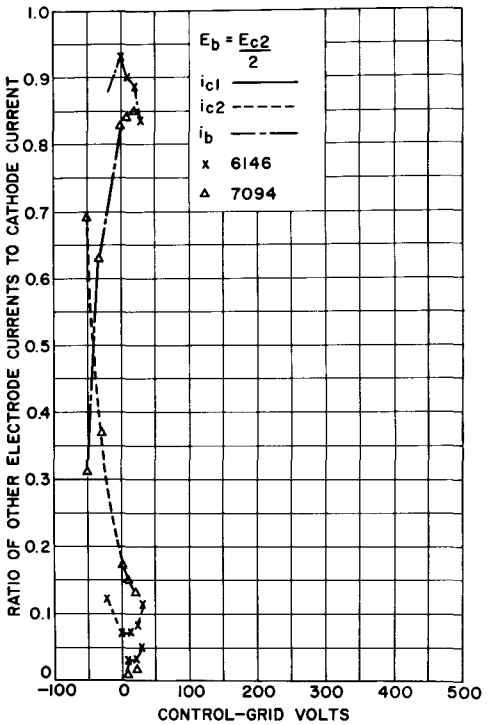


Figure 14. Current Division in Typical Tetrodes

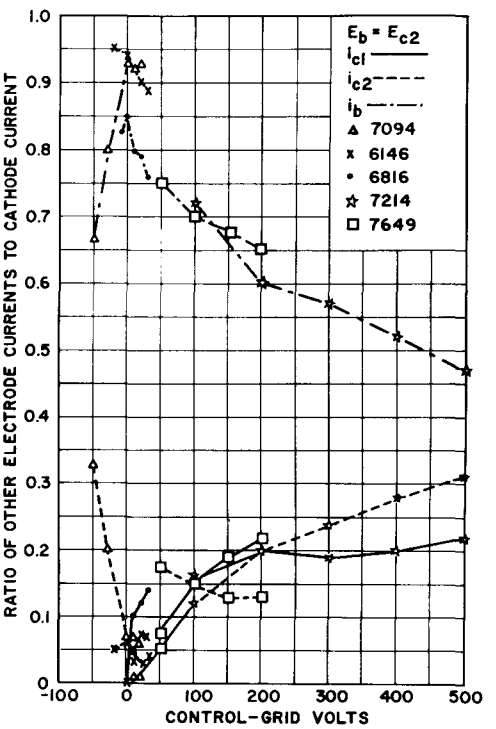


Figure 15. Current Division in Typical Tetrodes

The energy of the electrons as they pass through the screen-grid wires must be such that during the period

of minimum plate voltage they are decelerated to zero energy in some plane closer to the plate than to the screen grid. The space charge formed will then suppress secondary emission from the plate. K. R. Spangenberg[8] gives an excellent presentation of the effect of space charge on potential in the screen-grid and plate region of tetrodes and explains the hysteresis curve of tetrode plate and screen-grid current characteristics.

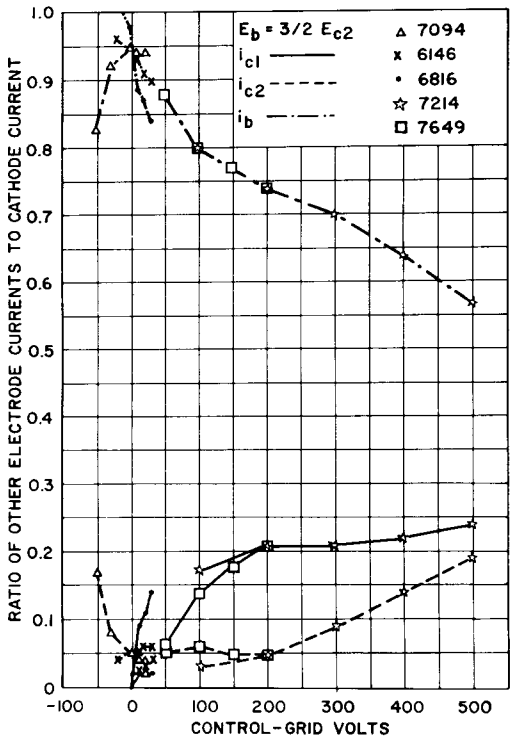


Figure 16. Current Division in Typical Tetrodes

Determination of Cathode Current

The relationship between electrode geometry and cathode current has been derived by a number of authors with the addition of quasi-correction factors for geometries which do not conform to the strict mathematics. For this reason, the usually applied equations are merely outlined in this section. The designer should refer to the references for detail and use the material presented here as reference only.

For Diodes:

$$i_s = \frac{2.33 \times 10^{-6} A}{d_{gk}^2}\ eb^{3/2} = G\,eb^{3/2}$$

where

i_s = cathode current - amperes

G = perveance

A = emitting area of cathode (sq. in.)

d_{gk} = grid-cathode spacing (in.)

e_b = plate potential volts

For Triodes:

$$i_s = \frac{2.33 \times 10^{-6} A}{d_{gk}^2}\left(\frac{e_{c_1} + \dfrac{e_b}{\mu_p}}{1 + \dfrac{1}{\mu_p} + \dfrac{4}{3\mu_p}\dfrac{d_{gp}}{d_{gk}}}\right)$$

where

e_{c_1} = control-grid voltage

μ_p = plate-to-grid amplification factor for triode

d_{gp} = grid-to-plate spacing (in.)

For Tetrodes:

$$i_s = \frac{2.33 \times 10^{-6} A}{d_{gk}^2}\left(\frac{e_{c_1} + \dfrac{e_{c_2}}{\mu_{c_2}} + \dfrac{e_b}{\mu_p}}{1 + \dfrac{1}{\mu_{c_2}} + \dfrac{4}{3\mu_{c_2}}\dfrac{d_{gs}}{d_{gk}}}\right)^{3/2}$$

where

e_{c_2} = screen-grid voltage

μ_{c_2} = screen-grid-to-control-grid mu factor

μ_p = plate-to-control-grid mu factor for tetrode

d_{gs} = control-grid-to-screen-grid spacing (in.)

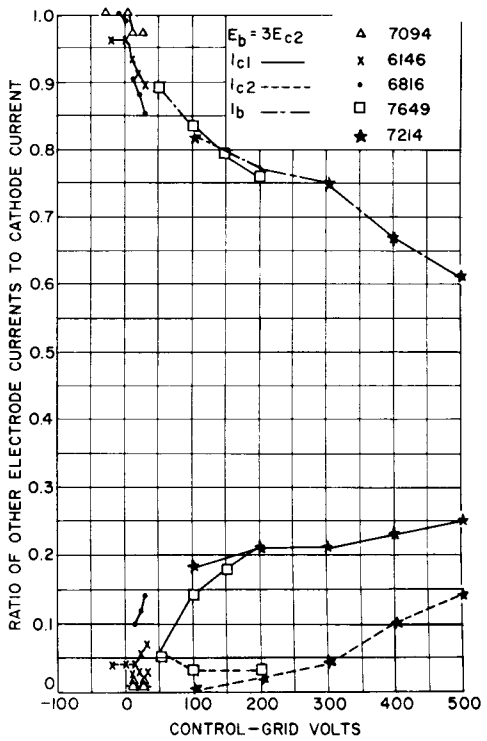


Figure 17. Current Division in Typical Tetrodes

These equations are based on the following assumptions: (1) The cathode is capable of emitting more electrons than are drawn to the plate (space-charge-limited). (2) The cathode and plate are parallel planes which are large enough to render end effects negligible.

(3) Plate and cathode surfaces are equipotential. (4) Electrons are emitted with zero initial velocity.

Correction factors have been derived by G. R. Partridge[9] for physical structures which do not fulfill the above assumed conditions.

Thermionic Emission[10]

Cathode emission must be high enough to provide space-charge-limited condition at the desired operating current level. The emission is related to cathode parameters as follows:

$$i_s = BAT^2 e^{-E_w/E_T}$$

where

B = Dushmann's constant

A = emitting area of cathode (m^2)

T = absolute temperature of cathode (deg K)

E_w = work function of emitter (ev)

E_T = T/1160, electron-volt equivalent of temperature

For commonly used emitters, the constants involved are approximately as shown in Table VI.

Table VI

Emitter	B amp/m^2 $^oK^2$	E_w ev
Tungsten	60.2	4.52
Thoriated Tungsten	3.0	2.63
Oxide Coated	0.01	1.0

The theoretical prediction of emission capability does not include the effects of cathode deterioration during life. Because good life expectancy is a design requirement, the cathodes in power tubes are rarely operated at their maximum capability. A cathode-current density of 100 to 200 milliamperes per square centimeter is used in practical designs for continuous-current operation to provide the desired margin of safety. The available current density for pulse operation is shown in Fig. 18. The use of these "rule-of-thumb" cathode-current densities in conjunction with the estimate of current division allows the designer to calculate the cathode area required to provide the necessary plate current.

Amplification of Mu Factor

The cathode-current equations require the determination of an amplification factor (for triodes) or mu factor (for tetrodes and pentodes) to complete the calculation. These factors are best defined by their mathematical use in the cathode-current equation:

$$i_s = G\left(e_{c_1} + \frac{e_{c_2}}{\mu_{c_2}} + \frac{e_b}{\mu_p}\right)^{\eta}$$

where

μ = amplification or mu factor

i_s = cathode current in amperes

G = perveance

e_{c_1} = control-grid voltage

e_{c_2} = screen-grid voltage

e_b = plate voltage

For a tetrode,

$$\mu_{c_2} = -\left(\frac{\delta e_{c_2}}{\delta e_{c_1}}\right)_{i_b, e_b}$$

$$\mu_p = -\left(\frac{\delta e_b}{\delta e_{c_1}}\right)_{i_b, e_{c_2}} = -\left(\frac{\delta e_b}{\delta e_{c_2}}\right)_{i_b, e_{c_1}} \left(\frac{\delta e_{c_2}}{\delta e_{c_1}}\right)_{i_b, e_b}$$

$$= \mu_{pc_2} \mu_{c_2}$$

For a triode,

$$\mu_p = -\left(\frac{\delta e_b}{\delta e_{c_1}}\right)_{c_b}$$

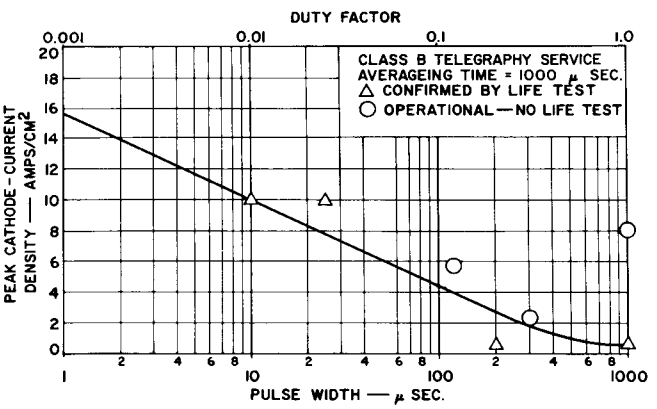


Figure 18. Present State-of-the-Art Pulse Current Density from Oxide Coated Cathodes

The method of measuring this factor is as follows: A tube is set up in a test circuit at initial control-grid and screen-grid or plate potentials Care must be taken to avoid excessive electrode dissipation if dc potentials are applied. The grid potential is changed slightly and plate current is brought back to its initial value by adjusting screen-grid or plate voltages. The amplification or mu factor is the ratio of the change in potential of the screen grid, suppressor grid, or plate to the change in control-grid potential at constant current. The so-called triode mu of a tetrode is measured by holding plate voltage constant (corresponding to μ_{c2}), or by connecting the plate and screen grid. This latter measurement

does not yield a mathematically correct value for μ_{c2} but is accurate enough to control product quality.

Mu factors are calculated as follows:

For small changes in μ to adjust tube characteristics, etc.

For tetrodes:

$$\mu_{c_2} = KN^2 Dd_{gs}$$

For triodes:

$$\mu_{p} = KN^2 Dd_{gp}$$

where

K = constant

N = control-grid turns per inch

D = control-grid wire diameter (inches)

d_{gs} = control-grid-to-screen-grid spacing (inches)

d_{gp} = control-grid-to-plate spacing (inches)

For quick calculation of μ for plate current above cutoff level, calculate μ from the empirical formula[11] as follows:

$$\mu = 2\pi Nd\left[0.2 + 6.8\,ND + 680\,(ND)^5\right]$$

This empirical formula is for μ_{c2} of a tetrode when d equals the control-grid-to-screen-grid spacing in inches and for μ_p of a triode when d equals the grid-to-plate spacing in inches.

The method by Vodges and Elder,[12] shown in Fig. 19, should also be used. Both answers should be corrected by the amount indicated on the empirical correction chart shown in Fig. 20. The average of these two corrected values is used as the final value.

For calculation of μ at plate-current, cutoff, see Ref. 13, Eqs. (16) and (17) in which

$$y = \frac{c}{2}$$

Mu μ is a measure of the influence of the screen grid or plate potential on plate current. For example, a low screen-grid mu results in large fluctuations of plate current with screen-grid voltage. If a tetrode has a screen-grid mu of one, the screen-grid voltage would be just as effective as control-grid voltage in controlling plate current. Thus, it would appear that tubes designed for screen-grid or screen-grid-and-plate modulation where screen-grid control of plate current is desired or high zero-bias plate current (linear amplifiers) where high zero-bias plate current is desired should have low mu. However, there are some real dangers to reducing the mu, as described below.

As an approximation:

$$\mu_{c_2} = KN^2 Dd_{gs}$$

where

μ_{c_2} = screen-grid-to-control-grid amplification factor

K = constant

N = effective turns per inch of the grid

D = control-grid lateral wire diameter

d_{gs} = control-grid-to-screen-grid spacing

To reduce mu, the lateral-wire diameter and control-grid-to-screen-grid spacing may only be reduced to a point where sufficient heat flow, beam formation, and mechanical stability are preserved. A further reduction in mu requires a reduction in effective turns per inch of the grid structure. This open-grid structure may result in an increase in the nonuniformity of the electric field at the cathode. This condition, called the "inselbildung" effect by many authors, causes remote cutoff and a nonlinear variation of plate current with control-grid voltage. These effects cannot be tolerated when linear modulation or simplification are desired. A good geometrical rule of thumb which yields a practical structure is TPI$\geq 3/2\ 1/d_{gk}$. Too many turns per inch result in excessive grid current because of unwanted blocking of the electron beam by the grid.

A high screen-grid mu in a tetrode usually indicates good linearity if it results from high number of turns per inch and large grid-to-cathode spacing. However, high-mu tubes usually do not have high zero-bias plate current because the grid-to-cathode spacing may be large and the effective accelerating potential of the screen grid, E_{c2}/μ_{c2}, is low. The power output of a linear amplifier is limited by this low zero-bias plate current because driving the control grid above zero loads the driver and introduces poor linearity unless the power output of the driver is much larger than actually required and is artificially swamped.

The inselbildung effect should always be considered as part of the mu problem. The reason for increasing the fineness of control-grid structures is to reduce inselbildung effects so that grid control with linearity may be preserved at high levels of plate current.

Transconductance

The transconductance may be defined and determined for a tetrode as follows:

$$g_m = \left(\frac{\delta i_b}{\delta e_{c_1}}\right)_{e_b,\, e_{c_2}} \approx \frac{KA}{d_{gk}^2}$$

$$\left(\frac{e_{c_1} + (e_{c_2}/\mu_{c_2})}{1+(1/\mu_{c_2}) + (4/3\mu_{c_2})(d_{gs}/d_{gk})}\right)^{1/2} = G\left(\frac{e_{c_1} + \dfrac{e_{c_2}}{\mu_{c_2}}}{1+\dfrac{1}{\mu_{c_2}} + \dfrac{4}{3\mu_{c_2}} + \dfrac{d_{gs}}{d_{gk}}}\right)^{1/2}$$

where

K = a constant

i_b = plate current

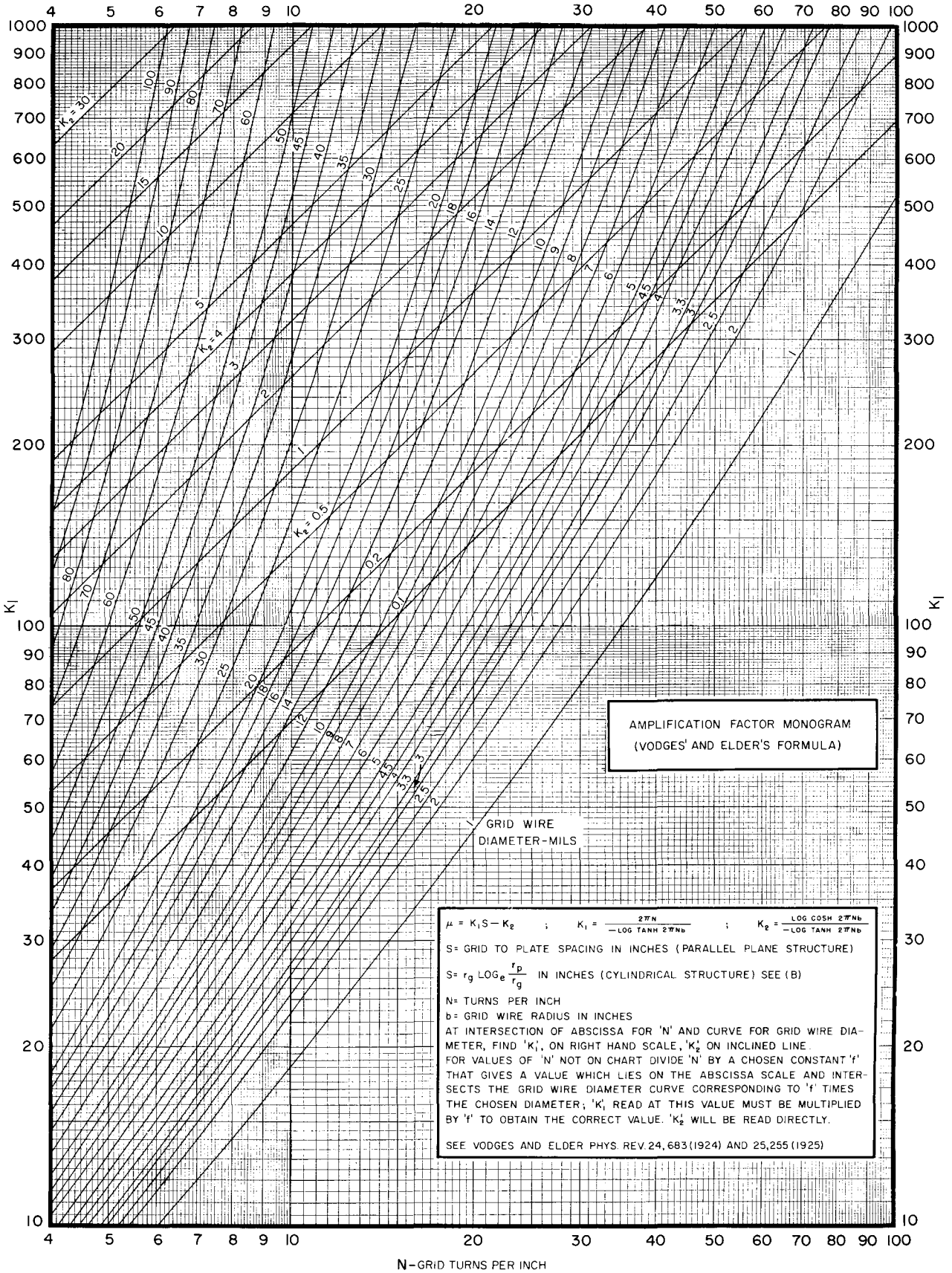

AMPLIFICATION FACTOR MONOGRAM
(VODGES' AND ELDER'S FORMULA)
GRID WIRE
DIAMETER-MILS
$\mu = K_1 S - K_2$; $K_1 = \frac{2\pi N}{-\text{LOG TANH } 2\pi Nb}$; $K_2 = \frac{\text{LOG COSH } 2\pi Nb}{-\text{LOG TANH } 2\pi Nb}$
S= GRID TO PLATE SPACING IN INCHES (PARALLEL PLANE STRUCTURE)
$S = r_g \text{ LOG}_e \frac{r_p}{r_g}$ IN INCHES (CYLINDRICAL STRUCTURE) SEE (B)
N= TURNS PER INCH
b= GRID WIRE RADIUS IN INCHES
AT INTERSECTION OF ABSCISSA FOR 'N' AND CURVE FOR GRID WIRE DIAMETER, FIND 'K_1', ON RIGHT HAND SCALE, 'K_2' ON INCLINED LINE.
FOR VALUES OF 'N' NOT ON CHART DIVIDE 'N' BY A CHOSEN CONSTANT 'f' THAT GIVES A VALUE WHICH LIES ON THE ABSCISSA SCALE AND INTERSECTS THE GRID WIRE DIAMETER CURVE CORRESPONDING TO 'f' TIMES THE CHOSEN DIAMETER; 'K_1' READ AT THIS VALUE MUST BE MULTIPLIED BY 'f' TO OBTAIN THE CORRECT VALUE. 'K_2' WILL BE READ DIRECTLY.
SEE VODGES AND ELDER PHYS. REV. 24, 683 (1924) AND 25,255 (1925)
N-GRID TURNS PER INCH
K_1

Figure 19. Penetration Factor (Kleijnen)

e_{c_1} = control-grid voltage
e_{c_2} = screen-grid voltage
e_b = plate voltage
d_{gk} = control-grid-to-cathode spacing
d_{gs} = control-grid-to-screen-grid spacing
μ_{c_2} = control-grid-to-screen-grid mu factor
A = active cathode area
G = perveance

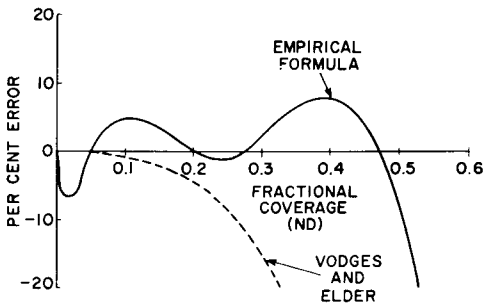


Figure 20. Correction Factors for Determination of Amplification Factor

The above equation is derived from the cathode-current equations by differentiating with respect to control-grid voltage.

For a grid-driven amplifier, power gain PG is affected by transconductance as follows:

$$PG = \frac{UPO}{DPO} = \frac{1/2\,\eta\, E_p I_p}{\hat{e}_{c_1} I_{c_1} - E_{c_1} I_{c_1} + \frac{(\hat{e}_{c_1} - E_{c_1})^2}{2r}} = \frac{\eta R \hat{i}_b^{\,2}}{2m^2\left(E_g I_{c_1} + \frac{E_g^{\,2}}{2r}\right)}$$

when

$$E_g I_{c_1} << \frac{E_g^{\,2}}{2r}$$

which is usually true for frequencies above 100 Mc because of the effects of reactance induced input admittance and transit time induced input admittance.[14, 15]

$$PG = \frac{\eta R r}{m^2} g m^2$$

where

UPO = useful power output
DPO = driver power output
η = output circuit efficiency
E_p = amplitude of plate-voltage fluctuation
I_p = fundamental component of plate current
E_g = amplitude of control-grid-voltage fluctuation $\hat{e}_{c_1} - E_{c_1}$
I_{c_1} = average control-grid current
r = shunt resistor representing losses in control-grid circuit (from leakage, skin resistance, contact resistance, dielectric heating, radiation loss, etc.)
R = plate-load impedance = E_p/I_p
m = $\hat{i}_b/I_p$ as a function of conduction angle
$\hat{e}_{c_1}$ = peak control-grid voltage
E_{c_1} = dc control-grid bias voltage (when bias is negative, $-E_{c_1}$ is positive)
$\hat{i}_b$ = maximum instantaneous plate current

Power gain is proportional to a power approaching the square of the transconductance. At frequencies above 800 Mc, this dependence is not valid because of output transit time effects and current division variations. Transconductance may be increased as the square of the control-grid-to-cathode spacing reduction as long as current division does not change. Consideration of the parameter m leads to the conclusion that power gain increases as the plate current conduction angle increases.

Perveance

The perveance of a tetrode may be expressed as:

$$G = \frac{KA}{d_{gk}^2}$$

where

K = a constant
A = the cathode area
d_{gk} = control-grid-to-cathode spacing

Perveance appears in the cathode-current equation as follows:

$$i_s = \frac{KA}{d_{gk}^2}\left(\frac{e_{c_1} + \frac{e_{c_2}}{\mu_{c_2}}}{1 + \frac{1}{\mu_{c_2}} + \frac{4}{3\mu_{c_2}}\frac{d_{gs}}{d_{gk}}}\right)^{3/2} = G\left(\frac{e_{c_1} + \frac{e_{c_2}}{\mu_{c_2}}}{1 + \frac{1}{\mu_{c_2}} + \frac{4}{3\mu_{c_2}}\frac{d_{gs}}{d_{gk}}}\right)^{3/2}$$

Plate current and transconductance are proportional to perveance; therefore, (1) Power gain is proportional to the square of perveance providing the frequency is between 100 and 800 megacycles. (2) Zero-bias plate current for high linear-amplifier power output is pro-

portional to perveance. (3) Higher circuit efficiency, bandwidth, and frequency of lumped-circuit operation can be achieved by increasing perveance, as shown by combining the following equations:

$$G = KI_p, \frac{E_p}{I_p} = \frac{Ql}{\omega_c}, \quad \eta = 1 - \frac{Ql}{Qu}, \quad Ql = \frac{f}{\Delta f}$$

To arrive at the following relationships for η, Δf, and f:

$$\eta = 1 - \frac{\omega C E_p}{GKQu}$$

$$\Delta f = \frac{KG}{2\pi C E_p}$$

$$f = \frac{KQu(1-\eta)G}{2\pi E_p(C_o + C_t)} \quad \text{in limit where } C_t \longrightarrow 0$$

where

K = constant
Ql = loaded Q
Qu = unloaded Q
$\omega = 2\pi f$
$C = C_o + C_t$
C_o = output capacitance of tube
C_t = tuning capacitance
f = frequency
Δf = bandwidth

DEFINITIONS OF SYMBOLS USED

A = Emitting area of cathode

a = A constant

B = Dushmann's constant

C = Capacitance - farads

C_o = Output capacitance of tube

C_t = Circuit and tuning capacitance

D = Grid lateral-wire diameter - inches

DP = Driving power - watts

DPO = Driver power output - watts

d = Distance between tube electrodes - inches

d_{gs} = Control-grid-to-screen-grid spacing - inches

d_{gk} = Control-grid-to-cathode spacing - inches

d_{gp} = Control-grid-to-plate spacing - inches

d_{sk} = Screen-grid-to-cathode spacing - inches

E = Amplitude of sinusoidal voltage - volts

E_b = DC plate-to-cathode voltage - volts

E_{bb} = DC plate-supply voltage - volts

E_{c_1} = DC control-grid-to-cathode voltage - volts

E_{c_2} = DC screen-grid-to-cathode voltage - volts

E_g = Amplitude of control-grid-voltage fluctuation - volts

E_p = Amplitude of plate-voltage fluctuation - volts

E_T = Electron-volt equivalent of temperature-electron volts

E_w = Work function - electron volts

e = Instantaneous voltage - volts

e' = Effective grid voltage - volts

e_{c_1} = Instantaneous control-grid voltage - volts

e_{c_1} = Peak control-grid voltage - volts

e_{c_2} = Instantaneous screen-grid voltage - volts

e_b = Instantaneous plate voltage - volts

f = Frequency - cps

Δf = Bandwidth - cps

G = Perveance or conductance

gm = Grid-to-plate transconductance - mhos

I_b = DC plate current - amperes

I_{c_1} = DC control-grid current - amperes

I_{c_2} = DC screen-grid current - amperes

I_p = Fundamental or higher-harmonic component of plate current - amperes

i_b = Instantaneous plate current - amperes

i_{c_1} = Instantaneous control-grid current - amperes

i_{c_2} = Instantaneous screen-grid current - amperes

i_s = Instantaneous cathode current - amperes

J = Cathode-current surface density

K = A constant

L = Inductance - henries

l = Length

M = Mutual inductance - henries

$m = \hat{i}_b/I_p$ (a function of conduction angle)

N = Turns per inch of grid lateral wire

PG = Power gain

PI = Power input - watts

Ql = Loaded Q

Qu = Unloaded Q

R = Resistance - ohms

R_{shunt} = Shunt resistance - ohms

R_{series} = Series resistance - ohms

r = Resistance - ohms

T = Absolute temperature - deg K

TPO = Tube power output - watts

t = Time - seconds

UPO = Useful power output - watts

W = Energy - joules

X = Reactance - ohms

Y = Admittance - mhos

Z = Impedance - ohms

Z_o = Characteristic impedance - ohms

μ = Amplification or mu factor

μ_{pc2} = Plate-to-screen-grid mu factor

μ_{c2} = Screen-grid-to-control-grid mu factor

μ_p = Plate-to-control-grid amplification factor

η = Circuit efficiency

ω = Angular frequency - radians/sec.

ω_o = Resonant angular frequency - radians/sec.

$\delta = \omega/\omega_o$ at 1/2 power level of resonant curve

λ = Wavelength

$\beta = 2\pi/\lambda$

REFERENCES

1. Furst, U. R., "Harmonic Analysis of Overbiased Amplifiers," Electronics, Vol. 17, pp. 143-144, March 1944
2. Wagner, W. G., "Simplified Methods for Computing Performance of Transmitting Tubes," Proc. IRE, Vol. 25, pp. 47-77. January 1937
3. Code H1
4. Koros, L. L., "A Novel UHF High-Power-Amplifier System," RCA Review, Vol. XVI, No. 2, pp. 251-280, June 1955
5. Swift, J., "Disc-Seal Circuit Techniques," Journal of the British Institution of Radio Engineers, pp. 607-622, December 1955 and pp. 95-111, February 1956
6. Schade, O. H., "Beam Power Tubes," Proc. IRE, Vol. 26, No. 2, pp. 137-181
7. Spangenberg, K.R., Vacuum Tubes, pp. 169-200, McGraw-Hill, New York, 1948
8. Ibid., pp. 248-265
9. Partridge, G. R., "Factors Influencing the Perveance of Power-Output Triodes," Proc. IRE, pp. 87-94, January 1949
10. Millman, J., and S. Seely, Electronics, pp. 106 and 174, McGraw-Hill, New York, 1951
11. Herold, E. W., "Empirical Formula for Amplification Factor," Proc. IRE, Vol. 35, No. 5, p. 493, May 1947
12. Vodges and Elder, "Formulas for the Amplification Constant for Three-Element Tubes in Which the Diameter of Grid Wires is Large Compared to the Spacing," Phys. Rev., 24, p. 683, 1924 and 25, p. 255, 1925
13. Deb, S., "Amplification Factors and Mutual Conductance of a Beam Power Valve," Proceedings of the Institute of Electrical Engineers, pp. 469-474, July 1955
14. Llewellyn, Proc. IRE, Vol. 3, pp. 144-166, March 1944
15. Spangenberg, K. R., Vacuum Tubes

Large-Power-Tube Design, Processing, and Testing

W. P. Bennett

Lancaster

The design of a large power tube, like any other design problem, must start with a listing of the performance requirements which the tube is expected to satisfy. To determine these requirements, the designer must first establish the application and specific class of service of the tube.

APPLICATIONS AND SERVICE

The applications in which large power tubes are used can be divided into two major categories: control service and generator service. In control service, the tube is used as an "electronic valve" to control the flow of electrical energy from one circuit to another. In generator service, the tube is used with rf circuit elements to supply ac power, usually in the radio-frequency spectrum. The term generator is used loosely; the tube is not a true generator of electrical power, but actually converts dc power supplied to it into rf ac power. The primary objective is to supply large quantities of rf power as efficiently as possible.

These two major categories of tube applications may be further subdivided into classes of service. In control service, the tube may be used as an on-off switch, a rectifier permitting current flow only in one direction and blocking flow in the opposite direction, a regulator to maintain a constant value of voltage or current, or a variable circuit element (usually analogous to a resistor) to provide a voltage and/or current which varies in accordance with some specified signal. The following applications are typical of control service:

1. Switch tubes in a pulse modulator[1] (e.g., in a radar transmitter).

2. Modulator tubes in a plate-modulated broadcast transmitter[2].

3. Devices controlling the flow of power to a machine such as a particle accelerator or controlled atomic fusion device.

In generator service, the tube may be used as a self-excited oscillator or as an amplifier. It may be operated under continuous-wave, steady-state conditions, or it may be amplitude-modulated so that the output and input are varied in a prescribed fashion (generally to convey information or to program the power level for some special process). The following applications are typical of generator service:

1. Self-excited oscillators for the generation of rf power for induction heating of metals and dielectric heating of insulators[3].

2. Frequency-modulated broadcast and communications transmitters.

3. Amplitude-modulated broadcast transmitters for television and radio.

4. Pulsed radar transmitters.

5. Single-sideband communications transmitters[4].

6. RF power sources for particle accelerators[5].

7. Continuous-wave communications transmitters[6].

Once the application is known and the category and class of service have been established, the specific requirements for the tube should be listed. The following criteria are important in the design of tubes for use in both control and generator service:

(a) operating voltage and current (as well as the power to be controlled in control service);

(b) output power, current, and/or voltage as a function of time;

(c) power sensitivity or gain;

(d) operating frequency and bandwidth;

(e) environmental requirements (e.g., temperature range, altitude or pressure, radiation, shock and vibration, etc.);

(f) size, weight, and input-power limitations, if any;

(g) any specific conditions or limitations imposed by the equipment specifications (e.g., type of cooling, general class of tube, particular operating parameters such as maximum voltage or phase characteristics).

THE CONCEPTUAL DESIGN

When the fundamental performance criteria have been established, the next step is the selection of an over-all design. This task may range from the simple specifi-

cation of an existing design capable of meeting the performance requirements, through the rearrangement or modification of existing design concepts or components into a new combination, to the creation of a completely new design concept. When the development of new design concepts is required, the first phase is often carried out to demonstrate the feasibility of the new concepts, and the actual component development is later pursued as an engineering project.

The conceptual design also involves the selection and/or general outline of the major components of the device, such as the cathode, grid, plate, and vacuum envelope. This step often requires individual developmental treatment in the form of theoretical computations, model studies, and chemical-physical-mechanical development of an empirical nature to establish methods, processes, materials, and techniques necessary to the proper functioning of the component or concept in the over-all device.

DETERMINATION OF POWER REQUIREMENTS

The first step in the actual design process is usually the determination of the projected plate operating values for the tube. For control service these valves are generally specified in the objective requirements; for generator service it is necessary to compute them from the required power output to be delivered to the input terminals of the following device (antenna or load).

The value of power output specified to the tube designer is usually a minimum equipment or system requirement which must be achieved at all times and under all conditions. This objective power-output value must be adjusted upwards as necessary to take into account transmission losses, load-impedance mismatch, output-circuit losses, power-input variations, and normal variations from tube to tube.

1. Any transmission loss between the load at which the specified power is to be delivered and the output terminals of the tube-circuit power-generating unit must be considered; the specified power should be increased accordingly to determine the value which must be delivered at the output terminal of the tube-circuit unit to a matched transmission line.

2. If the load is not constant in impedance, the range of the load mismatch must be considered and the objective power increased to determine the average power required for an idealized matched load.

3. The output-circuit efficiency should be estimated (including circuit losses both external and internal to the tube vacuum envelope), and the power increased to obtain the value which must be generated at the electron interaction space within the tube.

4. Incidental variations in power input to the tube as a result of line-voltage and power-supply variations and equipment aging factors must also be considered, and the objective power increased accordingly.

5. An additional safety factor is usually included to accommodate normal variations from tube to tube, measurement uncertainties, and some degradation of tube performance with life.

When the minimum acceptable load power output has been adjusted to provide for circuit losses as described above, it is possible to specify the effective electronic power which must be developed at an idealized electron interaction space (a space short enough that it need not be regarded as a distributed element in the circuit wave space). Simplified computations[7] can then be used to calculate the plate volt-ampere operating conditions.

DETERMINATION OF PLATE VOLTAGE

The operating plate voltage is one of the first electrical parameters to be established. The value to be used for a particular tube depends on the intended mode of operation; continuous-wave, plate-pulsed, control-grid and/or screen-grid-pulsed, or plate-modulated. Table I shows upper limiting values for different modes based on actual operating experience. These values are not necessarily limits for future tube designs, but they represent the present "state of the art". The values listed are limits governed by the internal vacuum region of the tube; other limiting conditions are assumed to be idealized.

Additional plate-voltage limitations may be imposed on externally-circuited tubes by the possibility of external breakdown. In some cases, the use of gas under pressure (such as air or special insulating gases) or a dielectric fluid around the outside of the seal region helps to prevent external corona and arcing. In general, the limiting factors in control service are the gradients resulting from the plate-supply-voltage component; in generator service they are the gradients resulting from rf voltages.

These additional limitations resulting from external-breakdown considerations involve six major factors, as follows:

1. Frequency of operation. The rf voltage appearing at the external tube terminal is a function of the operating frequency. At low frequencies it is equal to the voltage inside the tube across the electron interaction space. As the frequency is increased, however, the voltage across the tube seal decreases to a minimum value at the "strapped" resonant frequency, and then starts to increase again. At frequencies 1.5 to 2 times the strapped resonant frequency, the rf voltage across the tube seal may be several times the value across the electron interaction space within the tube.

2. Clearances. As stated above, it is desirable to operate a tube at or below its strapped resonant frequency to keep the voltage across the output insulator low. To achieve a high strapped resonant frequency, however, it is necessary to reduce clearances between parts and to shorten insulators. This procedure is in direct opposition to the general rules for minimizing voltage gradients. In general, longer insulators and greater clearances between metal parts and between metal and insulation parts are conducive to lower gradients and are, therefore, capable

Table I

Upper Limiting Values of Plate Voltage

Type of Service	Maximum Pulse Duration (Microseconds)	Limiting Voltage (Kilovolts)	Cathode Type
Plate-Pulsed	10-15	50-60	Matrix Oxide
	100	35-40	Thoriated Tungsten
	2000	30-35	Thoriated Tungsten
Control-Grid- and/or Screen-Grid-Pulsed	2000	20-25	Thoriated Tungsten
	10-15	20-25	Matrix Oxide
Continuous-Wave	-	20	Thoriated Tungsten
Plate-Modulated (Carrier Condition)	-	15	Thoriated Tungsten

of withstanding higher operating voltages. As the required operating frequency increases, therefore, voltage gradients also increase and additional limitations are imposed on the operating plate voltage of externally-circuited tubes. (This limitation does not necessarily apply to integral-circuit vacuum tubes.)

3. Ceramic-to-Metal Seals. Almost all insulators used in high-power, high-frequency tubes are made of ceramic material having a high dielectric constant (about 9). Because this dielectric constant is substantially higher than that of air or vacuum, electric fields in the vicinity of the ceramic are distorted. As a result, any thin air space between metal and ceramic in the high-voltage region of a tube is subjected to extreme voltage gradients. External voltage breakdown usually occurs in these vulnerable areas of the tube, the metal-to-ceramic joints.

In the radial-compression seal shown in Fig. 1, there is always a thin air space between metal and ceramic which tapers down to the point of actual contact; potting compounds have been used with limited success to fill this small air gap. The metallized seal shown in Fig. 2 does not have this air-gap problem because the metallizing is intimately bonded to the ceramic, but the metallizing itself ends in a relatively sharp contour; this end contour provides a point of high-voltage stress because the metallizing is a conductor on which field lines terminate. In the improved seal shown in Fig. 3, the sharp edge of the metallizing is turned on a contoured ceramic to produce a relatively field-free region; even with this seal, however, high gradients exist at the ends of the ceramic under some conditions of rf voltage.

4. Internal Tube Arcing. Internal tube arcing may be caused by unusually high voltage gradients within the tube structure. Such gradients are caused by the shape and proximity of metal parts and insulators.

5. Type of Cathode. The permissible plate operating voltage for a particular tube is also affected by the type of cathode selected. The values given in Table I assume the use of the type of cathode best suited for the indicated class of service, based on previous experience. In general, the limiting voltages must be reduced if another type of cathode is used.

6. Operating Bandwidth. Permissible operating plate voltage may also be reduced by the electric or untuned bandwidth over which a tube is required to operate. Other things being equal, potential bandwidth generally increases as operating plate current is increased and voltage is decreased. To achieve a certain bandwidth value, therefore, it may be necessary to operate a tube below its full voltage and power-output capability.

The limitations described above apply more to generator service than to control service. Because the strapped resonant frequency is not a limiting consideration in control service, clearances and insulators may be made relatively large. As a result, uniform and minimum voltage gradients can be provided both inside and outside the tube. The limiting factors usually result from internal voltage breakdown in the electron discharge space. Experience indicates a limiting plate-supply voltage of about 40 to 50 kilovolts for large tubes using thoriated-tungsten cathodes in control service.

SELECTION OF THE CATHODE

Only two types of cathodes are currently being used in new large-power-tube designs: thoriated-tungsten[8, 9] and matrix-oxide. The pure tungsten cathodes of earlier tubes have been replaced by the carburized thoriated-tungsten cathode, which supplies more current for a longer expected lifetime at a much lower operating temperature. The thoriated-tungsten cathode is also more economical to operate because it requires less heating power. (See article "Tungsten, Thoriated-Tungsten, and Thoria Emitters" by W.E. Harbaugh in this book.)

The matrix-oxide cathode, which is relatively new

in the large-power-tube field, is a special adaptation of the oxide cathode in which the emission coating material is impregnated in a porous sintered-nickel matrix bonded to a supporting base metal. The matrix cathode is more suitable for use in large power tubes than the standard oxide-coated cathode for the following reasons:

1. The emission coating has better mechanical stability and adherence because the oxide is impregnated into a firmly bonded metal matrix.

2. The relatively high-conductivity metal matrix completely surrounds the individual particles of poor-conductivity oxide emission material and provides better thermal and electrical conductivity.

3. The emission surface is more resistant to damage from electron and ion bombardment or inadvertent arcing to the cathode. (See article "Oxide-Coated Emitters" by C. P. Hadley in this book.)

The type of cathode to be used in a particular tube depends upon the application requirements. The thoriated-tungsten cathode is recommended for all types of service in which the average power demand (duty factor) is substantial, i.e., for cw service and for pulse service in which the pulse duration exceeds 50 to 100 microseconds and/or the duty factor exceeds two per cent. The matrix-oxide cathode, which has a high peak-current capability in short-pulse service, is preferred when the pulse duration is less than 20 to 25 microseconds and the duty factor is less than one per cent. (See article "Fundamentals of Electron Emission" by L. S. Nergaard in this book.)

Table II shows the current state of the art in cathode design based on actual experience with both cathode types. Thoriated-tungsten cathodes are preferred in large power tubes for long-life service (in the order of 10,000 to 20,000 hours or more). Such service is based on operating temperatures adequate to supply current loadings of 3 to 4 amperes per square centimeter. For short-pulse service, however, thoriated-tungsten cathodes are less suitable than matrix-oxide cathodes because they require 5 to 10 times the heating power and supply only 1/3 to 1/4 the peak emission capability.

The one most important factor affecting the performance capabilities and life of a power tube is the cathode temperature. The accurate measurement of the cathode temperatures during design of the tube cannot be overemphasized. It is imperative that these measurements be made in an environment which simulates the final operating environment of the device in all respects, and that all the appropriate corrective factors be applied in determining the true operating temperature.

The operating plate voltage of a tube is also a factor in determining the most suitable cathode type, primarily in control service. For example, a hard-tube switch in a modulator is required to hold off high voltages for the major portion of the operating cycle. Voltage-holdoff capabilities between electrodes in vacuum depend upon the degree of the vacuum, the cleanliness and smoothness of the electrode surfaces, and the atomic structure of the metals used.

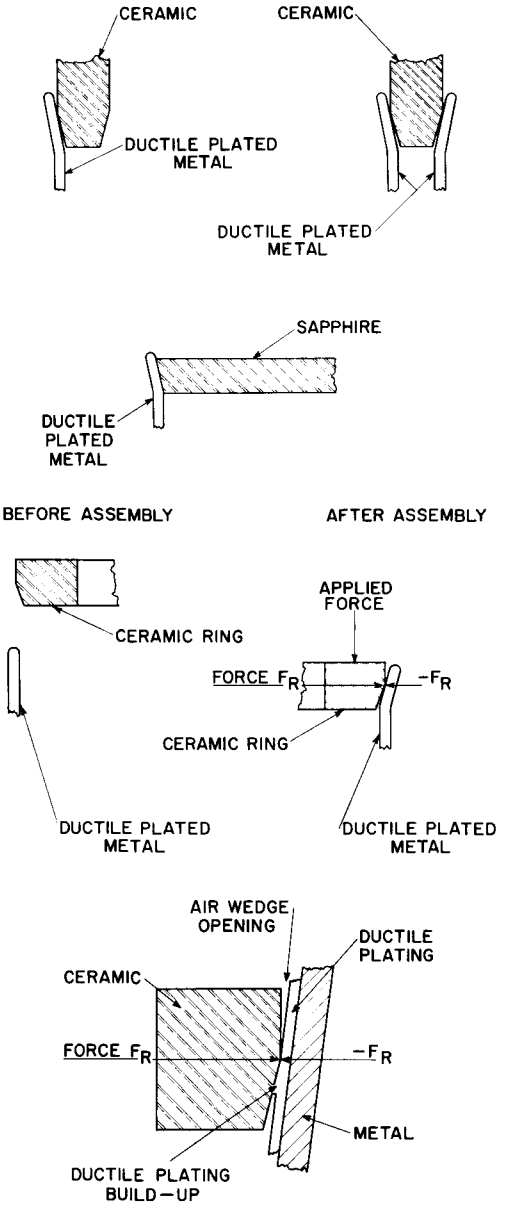


Figure 1. Radial Compression Seal

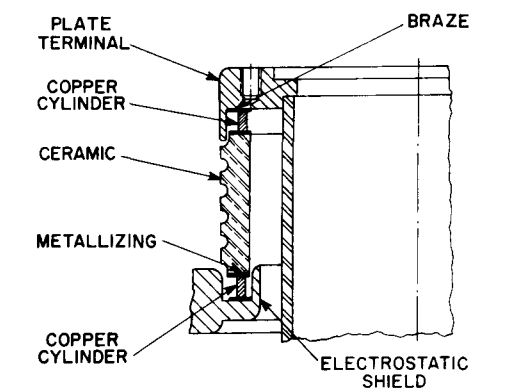


Figure 2. Metallized Seal

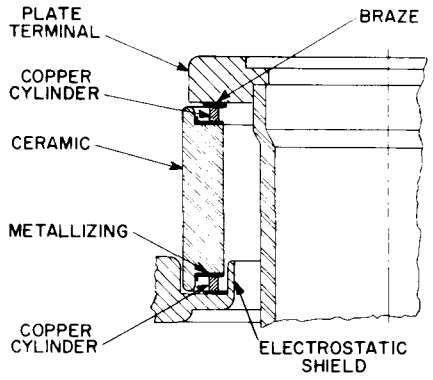


Figure 3. Improved Seal Using Contoured Ceramic

From the standpoint of voltage holdoff alone, pure tungsten would be the first choice for the cathode material. Because it has the lowest vapor pressure and is the most refractory of cathode materials, it can be processed most vigorously and thus contributes to the best vacuum conditions in the finished tube. In addition, it is a pure metal; therefore, it can be given a smooth finish and will not give off decomposition products during processing and life which may contaminate other tube electrodes and parts. Furthermore, its simple, easily maintained emitting surface is not easily destroyed by ion bombardment or contamination; contaminants which may be deposited on it can easily be evaporated by heating.

Thoriated tungsten is somewhat less desirable than pure tungsten for voltage holdoff because the thoria and carbon added to form the tungsten carbide layer evolve to some extent during tube processing and operation, and may contaminate other parts of the tube. In addition, the more complex carburized thoriated-tungsten emitting surface is more susceptible to ion bombardment and contamination, and is more difficult to restore to suitable condition after contamination.

The matrix-oxide cathode has the lowest voltage-holdoff rating because it is the most volatile cathode material. Because it evolves great quantities of gas during processing and to some extent during its operating life, it is not condusive to the best vacuum in the finished tube. It also has the disadvantages of a rough surface, and of substantial decomposition products which tend to contaminate other tube parts. However, it may be possible to enhance the performance capability of tubes employing matrix cathodes for high voltage use by employing ingenious methods for avoiding contamination of other tube parts during processing.

COMPUTATION OF THE REQUIRED CATHODE AREA

When the class of service, the required power output, the operating plate voltage, and the type of cathode have been determined, the required peak plate current can be computed. Allowance is then made for current absorbed by the control grid (and screen grid, if applicable) to determine the peak cathode-current demand. It is then a simple matter to determine the useful cathode emitting area required on the basis of the peak current rating of the specified cathode in amperes per square centimeter.

The first step is to determine under what conditions and at what point of the operating cycle the plate-current demand is the greatest in the particular class of service intended. The peak value required under the specified conditions is then determined. For switch-tube service, this value is usually the total value of load current which must be switched on and off. However, if fast pulse rise time is required, (as in short-pulse service), reactive charging currents can become appreciable. These currents must then be added vectorially to the resistive load-current components.

In class B rf power-amplifier service, the effective electronic power P which must be developed at the interaction space is given by

$$P = \frac{\hat{V}_p \hat{I}_p}{2} \qquad (1)$$

where V_p is the peak value of rf plate voltage developed across the interaction space and I_p is the peak value of the fundamental component of the electron space current to the plate.

Because the maximum current flows when the instantaneous plate voltage is at its lowest value, the average plate voltage $\bar{E}_b$ must be sufficiently greater than $\hat{V}_p$ to allow the plate-current pulse to pass to the plate. After the value of $\bar{E}_b$ is established, a value for $\hat{V}_p$ is assumed, and Eq. (1) is solved for $\hat{I}_p$. For class B service, the peak value of the current pulse is equal to twice the value of the fundamental component. For other classes of service, relationships can be developed for determining the value of the peak space current to the plate.

At this point, the designer must project the type of electron-gun configuration to be used in the new tube design. He can then estimate values for the space current absorbed by the control grid (and screen grid, if applicable), i. e., estimate the transmission efficiency of the gun. The peak cathode current $\hat{I}_k$ is equal to the peak plate current $\hat{I}_b$ divided by the gun efficiency. The useful emitting area required for the cathode can then be determined by dividing the value of $\hat{I}_k$ by the current rating (in amperes per square centimeter) for the selected cathode type and service application.

CATHODE CONFIGURATION AND SUPPORT STRUCTURE

After the plate operating voltage and current have been established and the required cathode area has been computed, the mechanical design of the tube begins. The first step is the selection of the mechanical configuration of the cathode and its supporting structure. Some of the typical arrangements in present-day large power tubes may be classified as follows:

(a) cylindrical, unipotential, indirectly-heated, matrix-oxide cathodes. Fig. 4 shows a cathode of this type used in the 7213 beam power tube. An emitting area of 40 square centimeters has been used in this type of cathode in a developmental tube.

(b) thoriated-tungsten, directly-heated, woven-wire

Table II

Cathode Design Data

Cathode Type	Pulse Duration	Max. Design Value for Peak Current (amp/cm^2)	Operating Temperature (K)	Design Life Expectancy (Hours)
Carburized Thoriated Tungsten	All Durations to cw	3 4	2000 2050	20,000 10,000
Matrix Oxide	To 15 μsec	15	1100	>2000

cathodes. Fig. 5 shows a cathode of this type presently being used in a developmental tube; this cathode has been used in European designs for some time.

(c) directly-heated, helical-wound, thoriated-tungsten wire cathodes. Fig. 6 shows this type of cathode as used in the 5762 power triode.

(d) directly-heated, cage-type, self-supported, thoriated-tungsten cathodes, formed of wires welded together in a "bud" at the top and supported on legs at the bottom. Fig. 7 shows this type of cathode as mounted in the 6166 beam power tube.

(e) hairpin-type, self-supported, directly-heated cathodes used in linear array structures. Fig. 8 shows this type of cathode as used in the 8D21 power tetrode.

(f) directly-heated, unitized cathode strands, suspended between clip-in grooves and mechanically held in place and kept in tension by a flexible pantograph or cantilever in which the tension is supplied by a spring device.

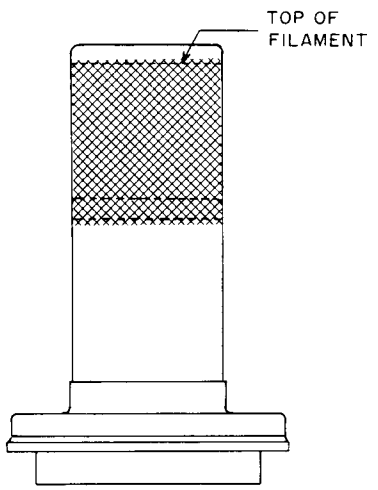


Figure 5. Thoriated-Tungsten, Directly-Heated, Woven-Wire Cathode Used in Developmental Tube

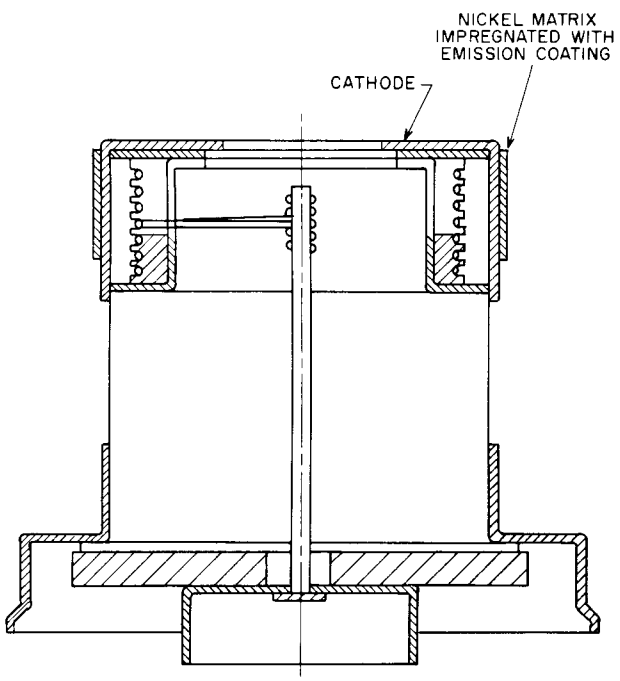


Figure 4. Cylindrical, Unipotential, Indirectly-Heated, Matrix-Oxide Cathode Used in RCA-7213

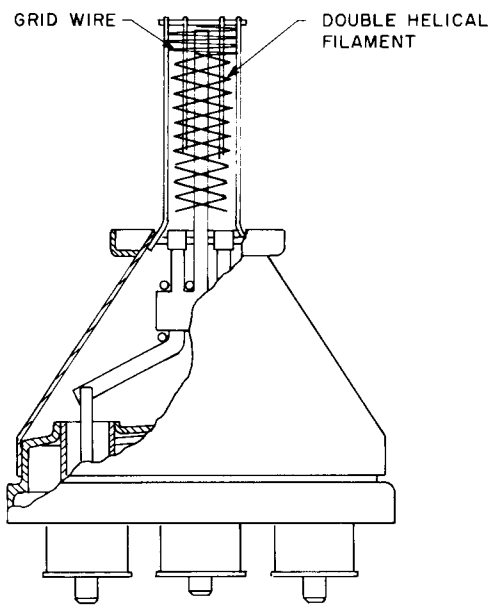


Figure 6. Directly-Heated, Helical-Wound, Thoriated-Tungsten Wire Cathode Used in RCA-5762

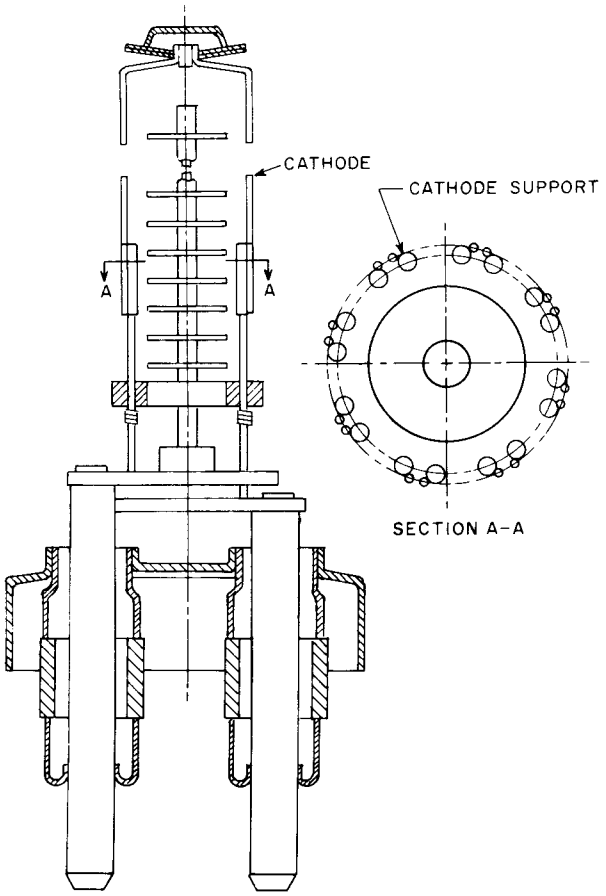


Figure 7. Directly-Heated, Cage-Type, Self-Supported, Thoriated-Tungsten Cathode Used in RCA-6166

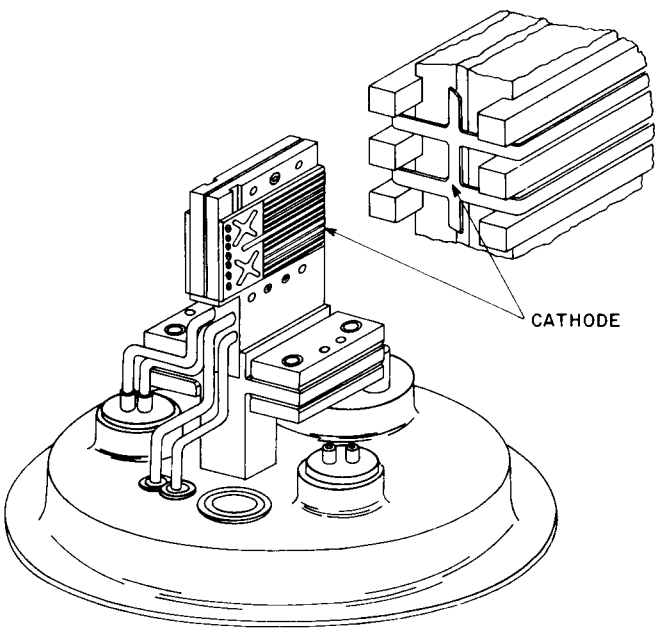


Figure 8. Hairpin-Type, Self-Supported, Directly-Heated Cathode Used in RCA-8D21

The complete mount structures shown in Figs. 4 through 8 can be scaled up or down in accordance with normal scaling rules, i.e., all dimensions, including control grids and screen grids, can be changed in linear proportion. For example, if the length and diameter of the cage-type structure were doubled (increasing the active area by a factor of 4), the diameter of the individual wire strands would also be doubled, as well as the clearance spacing between cathode strands and between the cathode and the control grid. When these types of structures are scaled upward to make larger tubes, therefore, the electron optical system (which involves primarily the control grid and the cathode) becomes increasingly coarser; the tube then requires proportionally higher driving voltages and shows proportionally poorer electron-beam transmission efficiency (higher grid-current absorption). Because driving power does not increase in direct proportion to the scaling factor, but at a rate several times greater, tube linearity characteristics also may deteriorate.

These disadvantages of upward power scaling are eliminated with the design approach used in the unitized cathode strand and electron optical system. Because similar units are added in parallel to make larger tubes, the wire size and clearance spacing between electrodes need not be scaled up in accordance with the desired power increase. Thus, the driving-power requirements need be increased only in linear relationship with the power output. For example, if four units are added in parallel, the active area is increased by a factor of 4, the driving power is increased by 4, and the power output is also increased by 4; as a result, the all-important power-gain factor remains constant.

The largest power tubes, therefore, and those having the greatest power-frequency products, use a multiplicity of electron optical systems in a parallel arrangement. The critical mechanical parameters of each unit are independent, but the units are coupled together electrically for operation as a single tube.

Unitized designs may use either thoriated-tungsten or matrix-oxide cathodes. The cathode is pre-fabricated in either case, and is also pre-processed in the case of thoriated-tungsten types. The ends of the filamentary-type thoriated-tungsten cathode strands are reduced in cross-sectional area to ensure a substantially uniform temperature over the entire length of the cathode between the terminal connections. (See article "Tungsten, Thoriated-Tungsten, and Thoria Emitters" by W. E. Harbaugh in this book.)

The heads of the filament strand must be appropriately shaped so that contact is made with the mooring system in such a way that the following objectives are achieved: (1) The filament strand must be accurately located and maintain its proper position throughout the life of the tube; (2) The contact between the filament strand and the mooring block must be adequate to carry the heating current and any rf currents without burning, pitting, or arcing at the contacts; (3) As the mooring system acts to compensate for filament expansion during heating or contraction during cooling, there must be no binding action nor torque transmitted to the filament strand which may cause it to "bow" or move out

of alignment; only direct in-line translational forces of tension should be applied to the filament strand.

The filament mooring system must also be designed so that it can carry the necessary heating current, but it must be sufficiently cooled so that it does not reach excessive temperature. Cooling is necessary to avoid loss of hardness of spring-tensioning members or excessive permanent deformation of the surfaces in contact with the filament strand or with other supporting members (e.g., push rods and rocker arms in the pantograph unit).

In addition, the mooring system must keep the filament strand under proper tension and accurately maintain it in its proper position throughout the life of the tube. It must compensate for expansion of the strand as it heats up to operating temperature, and must not be disturbed by on-off thermal cycling of the system. The tension in the system must be accurately controlled so that the strand is not loaded beyond its hot elastic limit nor relaxed to the point where loading at the contact between filament head and mooring block is insufficient to maintain an adequate current path for the heating currents.

A simple cantilever-type spring-tension mounting system is shown in Fig. 9; this system is useful only for short filament strands in which the total expansion during heating is relatively small. For longer filament strands with greater expansion, the more complex pantograph shown in Fig. 10 must be used.

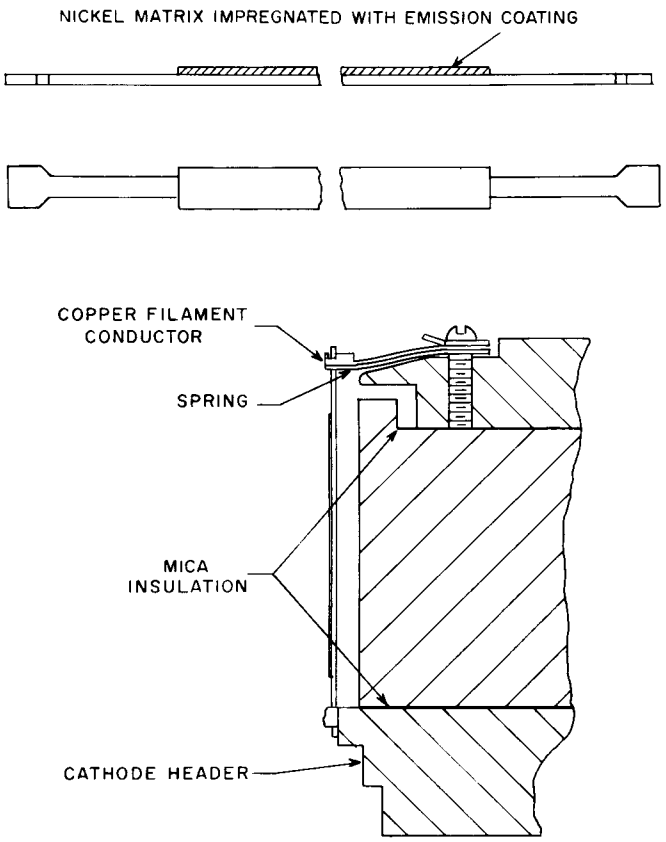


Figure 9. Cantilever-Type Spring-Tension Mounting System

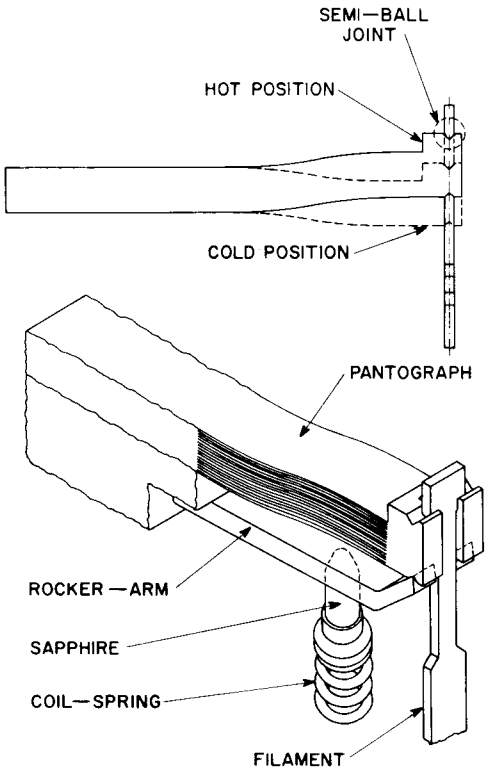


Figure 10. Pantograph-Type Mounting System

In the pantographic-type mounting system, the pantograph conducts the heating current to the filament strand, carries the actual mooring block, and positions the filament strand. It is made up of very thin (0.002-inch) chrome-copper-alloy laminations, which are gold-diffused together at their inner and outer peripheries. Individual pantographs or cantilevers may be clamped into the tube structure, or an assembly of pantographs may be made by diffusing disk-shaped laminations together into a disk pack. Radial cuts are then made in the laminated disk to form the individual support arms for the filament strands. The tensioning force is supplied by hardened tool-steel coil springs which are retained in tubular sleeves arranged in a cluster. Each spring is coupled to a ceramic or sapphire push-rod which, in turn, presses against a rocker arm which actuates the laminated pantograph. Thoriated-tungsten filament strands up to 12 inches long have been successfully mounted with this system.

TYPE AND CONFIGURATION OF TUBE

The next major step in the design of a large power tube involves the determination and/or selection of the type and general configuration of the tube. Power tubes may be classified in many ways. They may be (1) triodes,[10,11] tetrodes,[12,13,14,15] magnetrons,[16] klystrons,[17,18] or traveling-wave tubes;[19,20] (2) water- or liquid-cooled,[21] forced-air cooled,[22] or vapor-cooled;[23] (3) single-ended or double-ended;[10] (4) single-unit or twin or push-pull units within one envelope;[24] (5) designed for use with external circuits[25] or designed with integral-cavity circuits within their vacuum enclosure.[26]

The tube designer also has a choice of configurations for the tube. He may use a coaxial arrangement of

electrodes and supporting structures in which the electrodes are a series of concentric cylinders,[27,28] or he may use a planar arrangement with flat or nearly flat electrodes.[24] In the coaxial arrangement, the cathode may be the central member surrounded in turn by the other electrodes, or the plate may be the innermost electrode. An array of identical unit electron-optical systems (triode or tetrode) may be employed with individual unit electrode parts, or single "one-piece" cathodes and grids may be used.

The electron conduction system is the most important part of the tube. Other major areas of the design problem include the method of electrically connecting the elements of the electron conduction system to the appropriate circuit parts; the method of mechanically supporting the elements of the electron conduction system; provisions for cooling the elements, supporting structures, and internal circuit parts; and the design of an appropriate vacuum enclosure.

The elements of the electron conduction system are: the cathode, the cathode beam-forming electrodes (if used), the control grid (grid No. 1), the screen grid (grid No. 2), the post-screen-grid beam-forming electrodes or additional grids (if used), and the plate (anode). These elements will be discussed in greater detail after a brief consideration of the general effects of operating frequency on tube design. The intended frequency of operation affects the design of the electron conduction system, the interaction of the electron beam with the rf field, and the circuit components.

Frequency Effects in Control Service

In control applications, the signal or waveform may be periodic or aperiodic; it is best described as a series of transient conditions. The tube is used as a valve to control the flow of current, and therefore electrical energy, between appropriate parts of the circuit. It may provide a continuously varying state of the circuit, or successively different states with abrupt changes between states. In switching service, there are two states: "on" and "off".

The frequency-related effects are caused by the rate of change of current or energy as the circuit switches from one state to another. Any change can be described as a transient. A change can also be described in terms of, or resolved into, a combination of sinusoidal wave functions of appropriate amplitude and frequency. The transient phenomena can be related mathematically to the series of wave or frequency functions by means of Fourier Analysis.[29] In general, the more abrupt the transient change (the greater the rate of change with respect to time), the greater will be the value of the highest-frequency components of the Fourier Series. For the purpose of this discussion, quantitative determinations of the frequency components by Fourier Analysis will not be attempted; the treatment will be limited to consideration of the phenomena as simple transient conditions in a qualitative manner.

When large power tubes are used in control service, the electron can be considered to be without inertia. Therefore, the designer can disregard problems arising from the transit time of the electrons across the vacuum and the associated electron-interaction problems associated with flight time of the electrons. The two major items of concern then become the passive circuit properties of the tube (interelectrode capacitances and lead inductances) and the volt-ampere characteristics of the device.

Fig. 11 shows the simplified circuit of a triode tube used as a switch in a hard-tube modulator.[1,30] The switch tube turns the flow of current on and off in the circuit so that power is delivered to the load resistor in rectangular pulses. The waveshape of the current through, the voltage across, and the power delivered to the resistor is shown in Fig. 12.

When an idealized rectangular pulse having very rapid rise and fall times is required, tube performance may be limited by the interelectrode capacitances, or in some cases by the inductance or resistance of the tube leads, as well as by any capacitance shunting the load to ground. A more realistic equivalent circuit showing the tube capacitances and inductances is shown in Fig. 13. The added capacitances and inductances associated with the tube and load are not conducive to rapid pulse rise and fall times, and place additional requirements on the emission capabilities of the switch tube, the grid-driving supply, and the complexity of the circuit. In some cases, auxiliary circuit elements such as tail clippers are required to sharpen the pulse fall time. The aim of the tube designer, therefore, is to minimize both the tube lead inductances and the interelectrode capacitances.

In most cases, no special effort is needed to achieve values of lead inductance and resistance which are low enough to cause no appreciable detrimental effects. When very high currents and/or very short rise times are required, however, care should be taken in the design to assure that no problems will arise in this area.

The interelectrode capacitance values, on the other hand, are of sufficient magnitude to produce noticeable effects in a wide range of applications. The capacitance C_L in Fig. 13 consists of any capacitance shunting the load to ground, plus the capacitance from the space platform on which the modulator rests to ground. During the rise time of the pulse, C_L must be charged and C_{pk} (the tube plate-to-cathode capacitance) must be discharged; the necessary coulombs of charge must be supplied by current flowing through the switch tube. This current is in addition to the load current flowing through the resistor R_L, and, for fast rise times, may even exceed the load current.

In addition, the capacitance C_{gk} (tube grid-to-cathode capacitance) must be charged, and C_{gp} (tube grid-to-plate capacitance) must be discharged during the rise time; the coulombs of charge must be supplied by current from the grid driver. The faster the rise time, the greater the current demand.

During the pulse fall time, the capacitances are charged and discharged in reverse order. However, none of the charge is supplied by current flow through the switch tube. The fall time is dependent on the na-

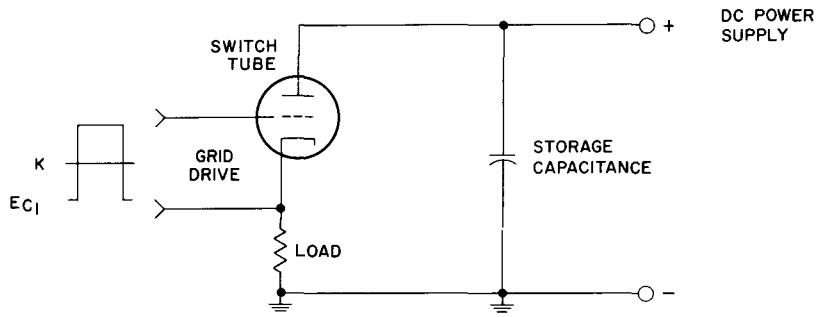


Figure 11. Simplified Circuit of a Triode Tube Used as a Switch in a Hard-Tube Modulator

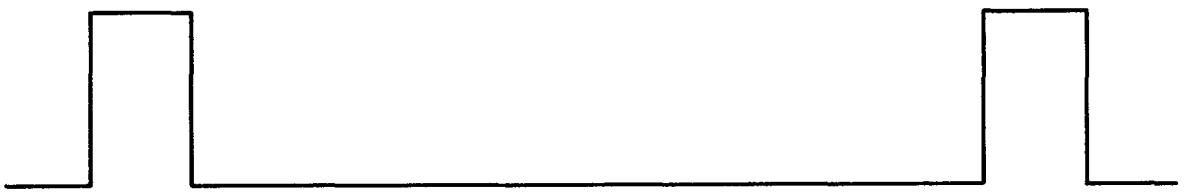


Figure 12. Waveshape of Current Through, Voltage Across, and Power Delivered to Resistor of Fig. 11

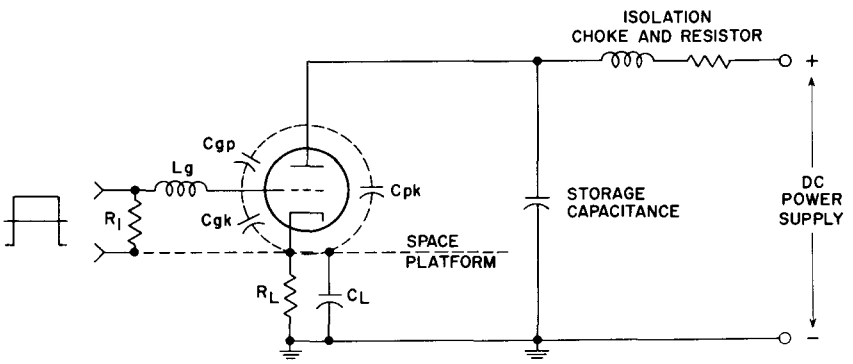


Figure 13. Equivalent Circuit for Hard-Tube Modulator

tural time constants of C_{gp}, C_{gk}, and R_1 in the grid-drive circuit and R_L, C_{pk}, and C_L in the load circuit. The impedance R_1 of the drive circuit may be reduced to sharpen the fall time in the grid-drive circuit, but at the expense of increased drive-power requirements. If the fall time controlled by the load circuit is too long, auxiliary circuits such as a tail clipper must be used to shunt the load at the end of the pulse and reduce the resistance at the circuit.

The tube designer can substantially reduce C_{gp} and minimize the effect of this capacitor by using a four-element electron conduction system in which a shield is placed between the plate and the grid. If this shield is returned directly to the cathode, the tube is called a shielded-grid triode; if the shield is returned to cathode through a large capacitor and is operated at a positive voltage with respect to the cathode, it is called a screen grid and the tube is called a tetrode. The shielded-grid triode is a favorite design for high-power switch-tube service because of the simplicity of its application and its rugged structure.

Switch-tube performance is also affected by the amount of pulse control-grid voltage required to change the tube from the "off" to the full "on" condition. The lower the voltage required, the lower will be the drive power required to overcome the limitations imposed by C_{gk} and C_{gp}. For this reason, electron conduction systems having high amplification factor (μ) and high transconductance (g_m) are preferred for switch tubes.

Frequency Effects in RF Generator Service

In rf service, the effects of increasingly higher operating frequency first appear as circuit problems, involving either the component parts of the tube as passive circuit elements or the choice of type and mode of

operation of the electron conduction system to provide optimum compatibility with the circuit aspects of tube operation. This situation is much the same as in the case of control service. As the frequency becomes higher to the point where one wavelength is approximately within one order of magnitude of the tube dimensions, the tube and its associated circuit components are no longer treated as lumped-constant elements, but as distributed-constant elements in the manner of transmission lines and waveguides. At the higher frequencies, therefore, additional boundary conditions are imposed by the tube dimensions, besides those imposed by mechanical and thermal considerations. Obviously, the larger and more powerful a tube must be, the lower will be the frequency limitations imposed by the tube dimensions.

As the frequency is increased, a point is reached where the electron can no longer be considered inertialess and the transit time of the electrons across the interelectrode spaces comes within an order of magnitude of the period of the operating frequency. The tube can then no longer be regarded as a valve for electromagnetic waves (i. e., a valve for controlling current as it flows in a metallic rod or resistor), but must be regarded as a region in which interaction between electrons and electromagnetic fields takes place in a controlled manner to produce high-frequency power. The study of the interaction of electrons with electromagnetic waves is extremely complicated. Much work has been done on the subject,[31] but there is still no method of analysis of electron-interaction phenomena for grid-controlled tubes which can be used by engineers to predict the effects of electron transit time beyond the limits of empirical experience.

Triodes

The simplest type of power-tube construction is the triode. In low-frequency service, the triode is usually connected as a grid-driven amplifier, as shown in Fig. 14. The merits of a high-power amplifier are based primarily on its power gain, efficiency, stability (freedom from self-excited oscillation and parasitic responses), and, in many applications, linearity and bandwidth.

For frequencies up to several megacycles, a triode connected as a grid-driven amplifier can perform very creditably, even though neutralization is required to balance out the feedback effects of the grid-plate capacitance C_{gp}. The merit of a particular triode type for low-frequency operation is governed almost completely by the quality of its electron conduction system, as indicated by the volt-ampere characteristics. (Electron conduction systems will be discussed in more detail later.)

Fig. 15 shows another popular method of operating power triodes, the push-pull circuit. Two tubes are connected in common input and output circuits in which the grids are driven 180 degrees out of phase and the plates also operate with out-of-phase voltages. The push-pull connection has many advantages, as follows:

1. Cross-neutralization may be employed.
2. Harmonic output is reduced.
3. Linearity may be improved.
4. The effective input and output capacitances are reduced.
5. The tube may work into a pure resistive load for audio and video amplifiers (no tank circuit needed for energy storage).
6. The outputs of two individual tubes can be combined.

In audio service, a single tube must be operated in class A service with attendant low efficiency (20 to 40 per cent). Tubes in push-pull, however, may be operated in class B service without degradation in performance and, in most cases, with improved linearity; thus efficiency can be increased to the normally high class B values (50 to 75 per cent).

As the operating frequency is increased to values above 10 to 20 megacycles, it becomes increasingly difficult, and finally impossible, to neutralize the grid-to-plate capacitance of a triode. At these higher frequencies, however, triodes can provide satisfactory performance in a cathode-drive (grounded-grid or grid-separation) circuit,[11, 27, 32] as shown in Fig. 16. Tubes intended for such service should be designed to minimize the plate-to-cathode capacitance C_{pk} (which becomes the feedback path). In addition, the grid should provide as complete a separation as possible between the plate and the cathode so that the output circuit is

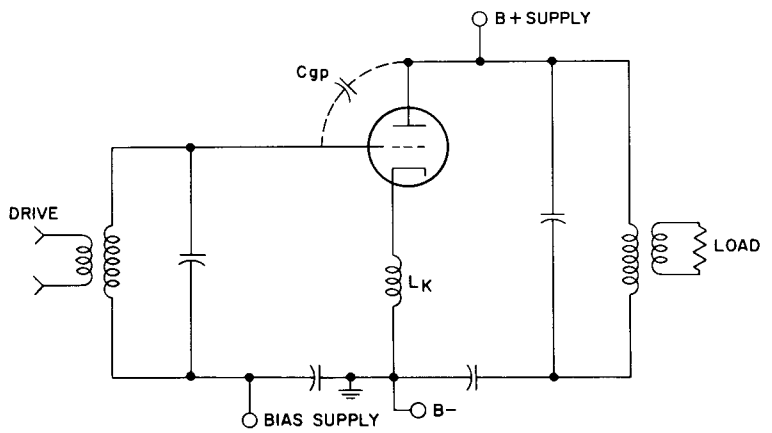


Figure 14. Simplified Circuit of a Triode Tube Connected as a Grid-Driven Amplifier

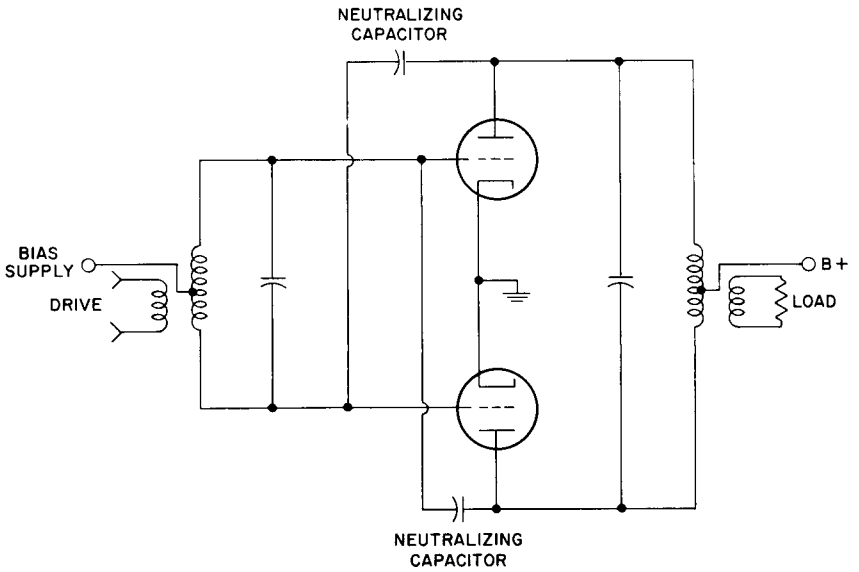


Figure 15. Push-Pull Triode Amplifier Circuit

isolated from the input circuit internally as well as in external connections.

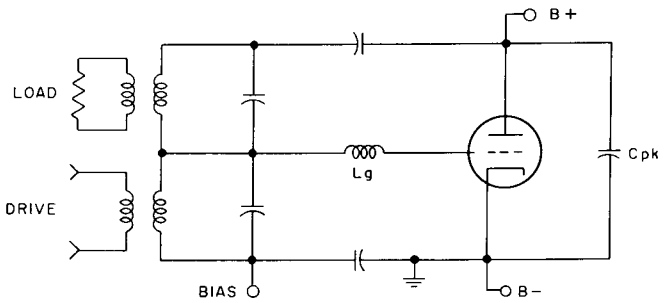


Figure 16. Triode Tube Connected in a Cathode-Drive (Grounded-grid or grid-separation) Circuit

The grid should also be supported on a continuous member which extends through the vacuum envelope as a solid shield. posts or support rods are not suitable grid-support structures for a cathode-driven triode because the relatively open grid structure represents substantial mutual inductive coupling between the input and output circuits which acts as another source of feedback. The right amounts of C_{pk} and L_g (mutual) in combination can neutralize each other, but usually only at one frequency or in a narrow range of frequencies.

One disadvantage of the cathode-drive amplifier is its relatively low gain compared to grid-drive types. Besides normal grid and circuit losses, there is considerable loss of driving power to the electron beam. When the cathode is driven negative to cause electrons to flow, the driver is placed in series with the plate supply. Thus, the power absorbed by the beam must be supplied by the driver. This power is equal to $(\hat{E}_k \hat{I}_p)/2$ where $\hat{E}_k$ is the peak value of the rf cathode-grid voltage and $\hat{I}_p$ is the peak value of the fundamental component of plate current.

The power absorbed by the beam need not be completely lost because it may be recovered in the output circuit. Furthermore, it represents a dissipative load which provides a built-in broadbanding of the input circuit. Unless regenerative feedback is employed, however, the upper limit of the gain of the cathode-driven stage is given approximately by $\hat{E}_p/\hat{E}_k$, where $\hat{E}_p$ is the peak value of the rf plate-to-grid voltage. For cathode-driven tubes, therefore, the electron conduction system would be designed to minimize the input or control voltage $\hat{E}_k$ required. High-mu, high-transconductance guns are desired.

Input-circuit losses may represent a substantial portion of the total drive-power requirements, particularly as the frequency becomes higher, and thus may further reduce the power gain. Because the circuit losses are proportional to the square of the drive voltage, it is almost mandatory to design the electron gun to minimize the peak input voltage required if power gain is to be maintained at a high value. It may even be desirable to do so at the expense of increased grid-current absorption for high-frequency tubes.

Tetrodes

The use of a tetrode-type tube reduces coupling between input and output in a cathode-drive circuit by the addition of a second grid member, the screen grid. However, the new circuit between control grid and screen grid must be handled properly or it may become a source of oscillations. Many straightforward engineering techniques are available to prevent tetrode oscillations.

The cathode-driven tetrode amplifier shown in Fig. 17 provides a potential reduction in driving power, and thus improved gain, as compared to triode circuits because less control-grid current is absorbed and less rf control-grid voltage is required. The screen grid is bypassed to the control grid for cathode-drive operation. Cathode-driven tetrodes (and triodes) can also be operated in push-pull to combine the outputs of two tubes.

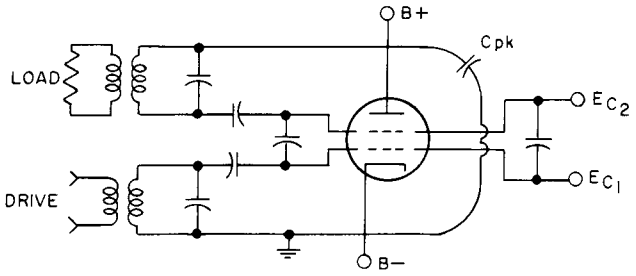


Figure 17. Cathode-Driven Tetrode Amplifier

Tetrodes provide even further advantages in grid-drive amplifiers[12] such as that shown in Fig. 18. The major advantages are as follows: (1) The shielding action of the screen grid reduces feedback effects. (2) The inherent high-gain characteristic of the grid-drive amplifier is retained. (3) Power gain is increased as a result of the effect of the screen grid in improving the control-grid volt-ampere characteristics.

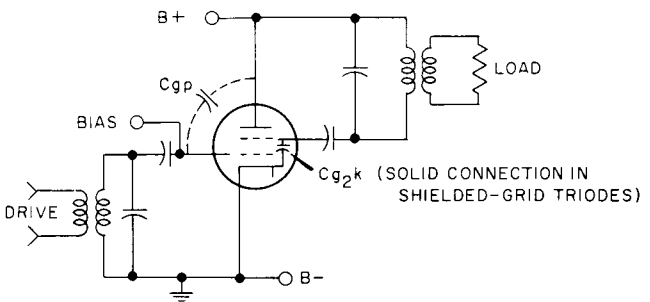


Figure 18. Grid-Driven Tetrode Amplifier

For frequencies up to the uhf region, an adaptation of the tetrode called a "shielded-grid triode" has been very successful. In this tube, the "screen grid" is connected directly to the cathode support structure within the vacuum envelope. This design provides good shielding, the inherent high gain of the grid-driven amplifier, and a tetrode-like plate characteristic. It has all the advantages of the tetrode, but does not require a screen-grid power supply and has no screen-grid dissipation problems. Because these tubes require rather high grid-driving voltages, however, they would result in rather high circuit losses if used in the ultra-high-frequency region.

Some tetrodes designed especially for uhf service incorporate a built-in screen-grid-to-cathode bypass circuit. These tubes are intended for use in grid-drive amplifiers, and use a high-performance gun that requires low grid-drive voltages.[12] Although these so-called "high-frequency features" also make the tubes more efficient at low frequencies, they are added to the tubes at extra cost. The uhf versions may be the best available technically for low-frequency applications, but they are not always economically competitive.

The grid-driven tetrode may, of course, also be operated in push-pull connection, as shown in Fig. 19, with the same advantages discussed previously for triodes.

Although present high-power data do not indicate any substantial degradation in performance due to transit-time effects, it is not advisable to extrapolate into frequencies at which no experience has been obtained with high-power tubes and say that transit-time problems will not arise. In future high-power tube designs, it may well be necessary to minimize transit time in the output regions by use of the "resnatron" principle.[13] This design, shown in Fig. 20, uses a high-voltage electrode adjacent to the control grid to accelerate electrons to high velocities before they enter the output space of the tube.

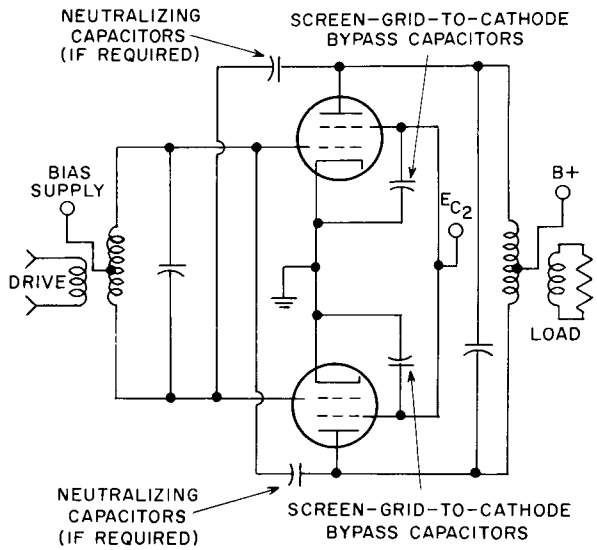


Figure 19. Push-Pull Grid-Driven Tetrode Amplifier

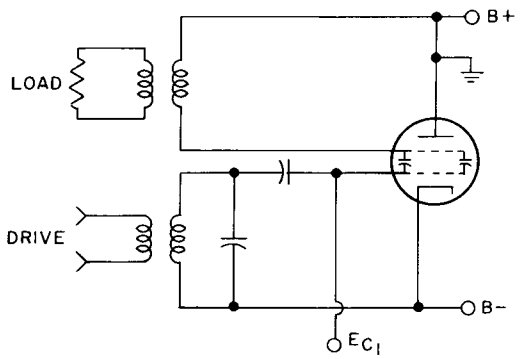


Figure 20. Schematic Diagram Illustrating Use of Resnatron Principle to Minimize Transit Time

Electron Conduction Systems

The electron conduction system is the most important part of the power tube. It involves the flow of electron space current through the vacuum region bounded by the various metallic electrodes. (See articles "Space Current Flow" by M. V. Hoover and "A New Approach to the Calculation of Electron-Tube Characteristics" by O. H. Schade, Sr., in this book.) Factors affecting the space-current flow include (a) the spacings between electrodes, (b) the shapes of the various electrodes, (c) the voltages applied to the electrodes, (d) the availability of electrons at the cathode, and (e) the total size (area) of the electrode array.

The function of the electron conduction system is to pass the maximum amount of electron current from cathode to plate with minimum electron absorption by the intervening grids (control electrodes). The positive plate potential required to draw electrons across the output space to the plate should be as low as possible. Similarly, the positive potential required at the control-grid electrode to cause electron-current flow and the negative potential required to cut off electron flow through the interaction space should be minimized.

Many applications also require that the volt-ampere characteristics of the tube exhibit a linear relationship between plate current and control-grid voltage. In other words, equal incremental changes in control-grid voltage should produce equal incremental changes in plate current (other parameters remaining constant) at any point in the dynamic operating range of the tube from cutoff to full plate current. Parameters which are important in specific applications can be optimized in particular tube designs at the expense of other, less important parameters.

The volt-ampere characteristics of a tube may also be affected by secondary emission of electrons from the control grid, screen grid, or plate, These secondary-emission effects can drastically influence the characteristics in certain regions of applied voltages.

Typical electrode configurations for a number of different types of electron conduction systems are shown in Figs. 21 through 29. In the triode configuration shown in Fig. 21, the directly-heated cathode consists of a relatively heavy wire wound into a double helix. The grid consists of lateral wires wound over or under the grid-support rods.

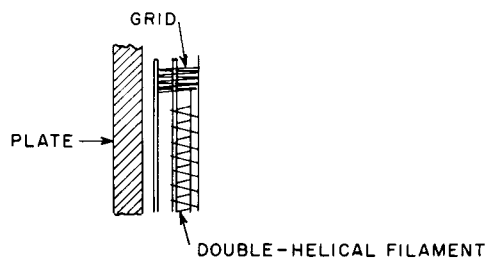


Figure 21. Triode Configuration Using Double-Helical Filament

The tetrode configuration shown in Fig. 22 employs an array of control-grid and screen-grid elements formed into a "squirrel cage" around a unipotential, continuous cathode surface; this type of configuration is used in the RCA types 6181[33] and 7213.[28] Fig. 23 shows a tetrode configuration with "squirrel-cage" grid structures as used in conventional large-power tubes such as the 6166.[15]

Fig. 24 shows part of the array of unit triode systems used in the electron-optical system of the 5831 beam triode. The triode units shown in Fig. 25 are used in the 6949 shielded-grid beam triode. Fig. 26 shows an array of unit triode systems presently being used in developmental high-mu tubes for uhf service.

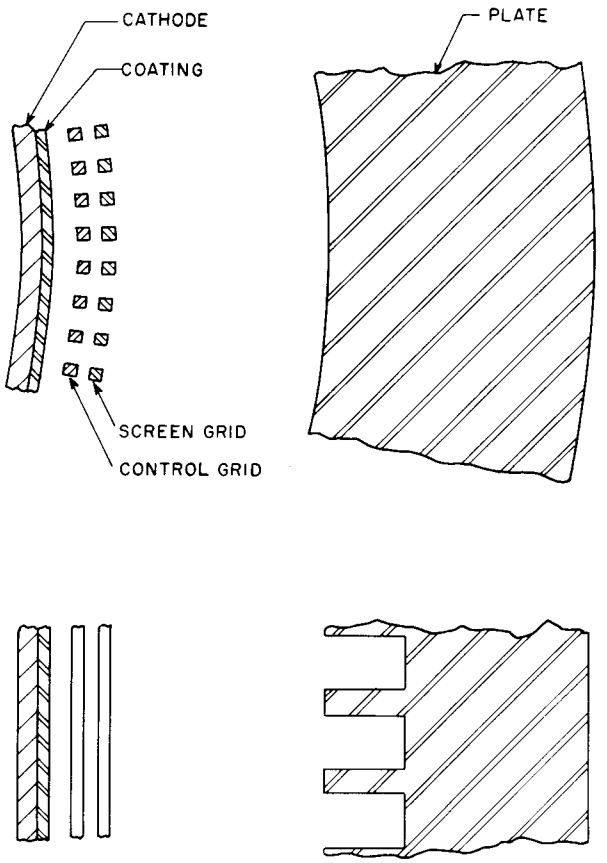


Figure 22. Tetrode Configuration Using Array of Control-Grid and Screen-Grid Elements around Continuous Cathode Surface

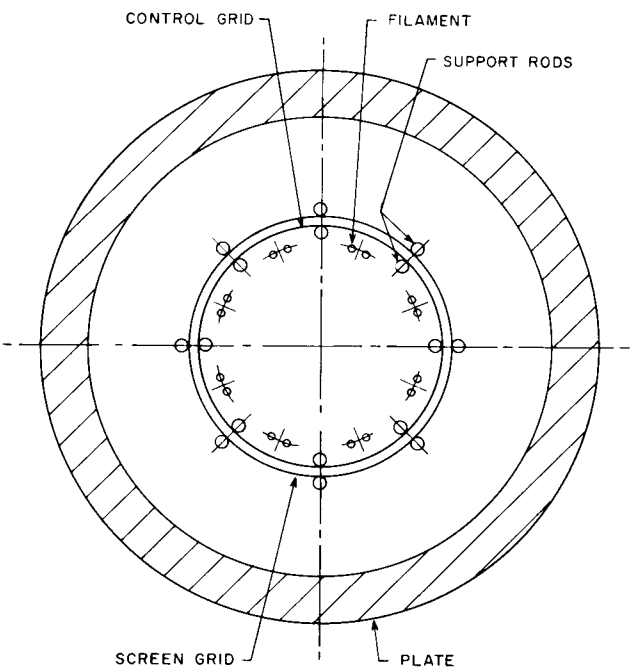


Figure 23. "Squirrel-Cage" Grid Construction Used in Conventional Large-Power Tetrodes

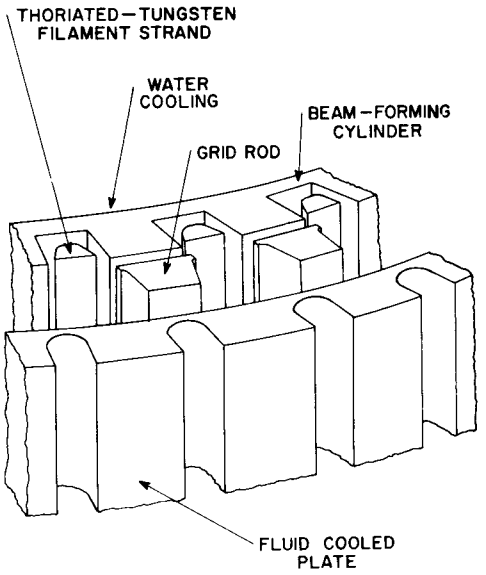


Figure 24. Unit Triode Systems Used in 5831 Beam Triode

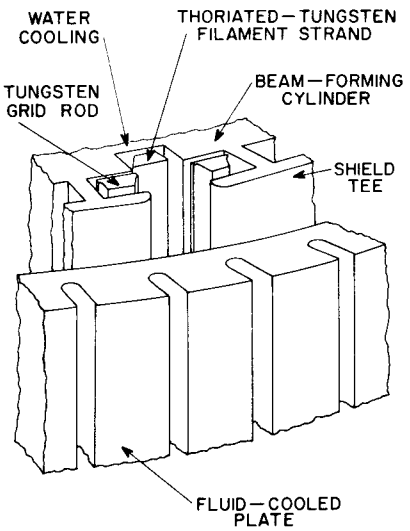


Figure 25. Triode Units Used in 6949 Shielded-Grid Beam Triode

Fig. 27 shows unit electron-control systems used in high-power beam tetrodes such as the 6448. A different beam tetrode configuration, shown in Fig. 28, is used in the 8D21. Fig. 29 shows the array of unit tetrode systems used in a tube designed according to the "resnatron" principle.

Beam-forming systems, such as the beam triode and the shielded-grid triode, generally provide superbly linear characteristics with extremely low grid-current absorption.[4] However, these systems require high grid-drive voltages. These structures are well suited to switch-tube and modulator applications, as well as to relatively low-frequency operation (to about 200 megacycles) where rf grid-drive voltage can be readily obtained without a premium on driving power. At higher frequencies (uhf), high drive voltages are difficult to achieve because of excessive input-circuit losses. Furthermore, if grounded-grid (cathode-drive) operation is employed, driving power is directly proportional to grid-drive voltage. Beam-forming systems require precise electrode shaping and dimensions, as well as precise positioning, particularly with regard to angular position of indexing (e.g., in a circular array of unit triodes).

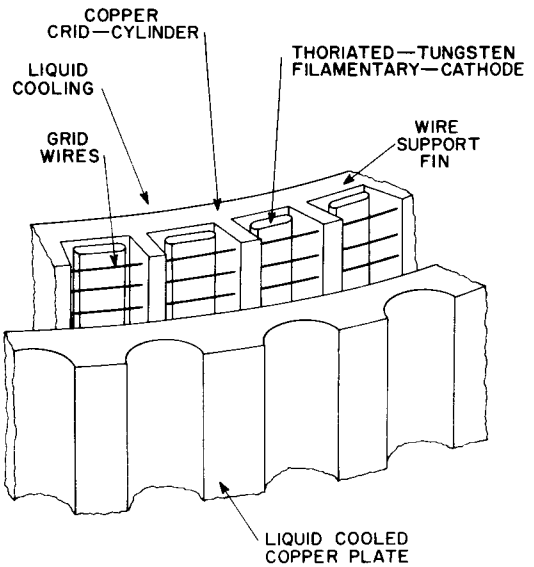


Figure 26. Unit Triode Systems Used in High-Mu Tubes for uhf Service

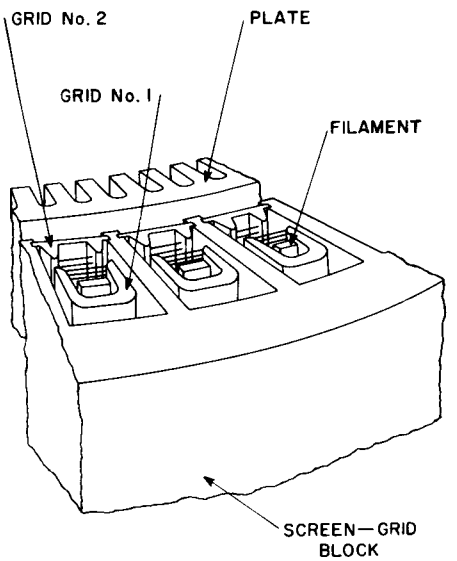


Figure 27. Unit Electron-Control Systems Used in High-Power Beam Tetrodes Such as the 6448

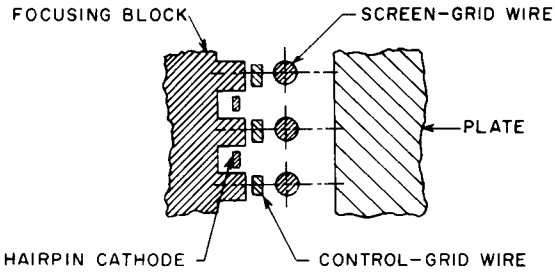


Figure 28. Beam Tetrode Configuration Used in 8D21

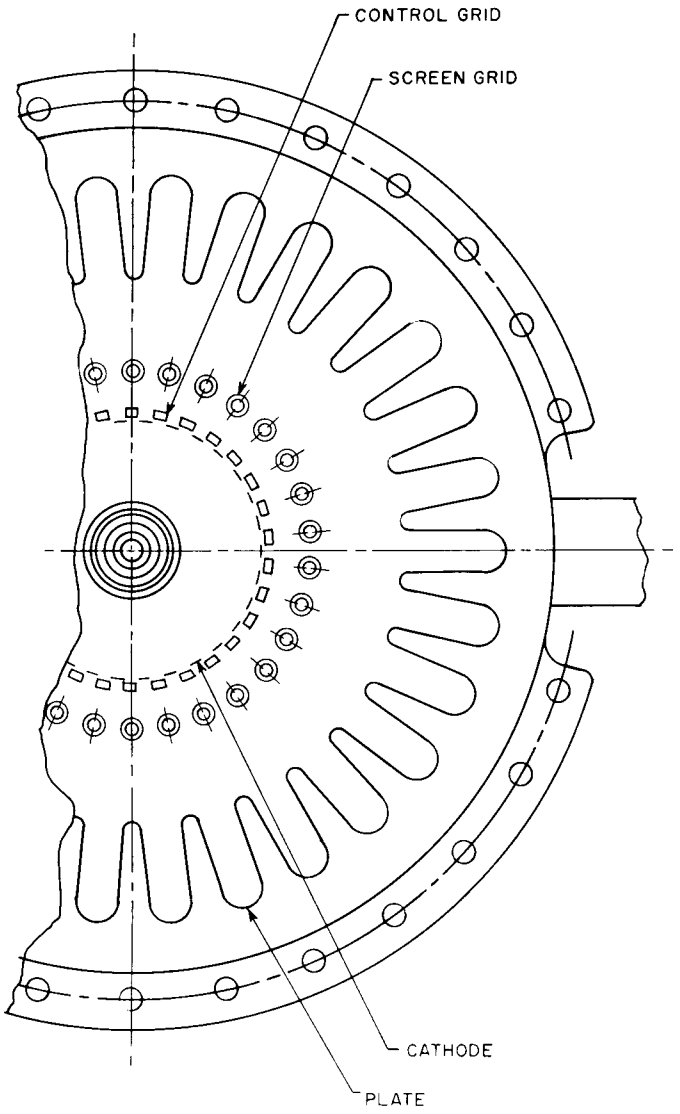


Figure 29. Unit Tetrode Systems Used in "Resnatron" Tube

Operation at higher frequencies usually requires the use of a "fine-grain" structure in which the cathode strip is spanned by small-diameter, closely-spaced control wires.[32] This arrangement requires considerably lower grid-drive voltages, but grid-current absorption is generally high. Angular position is not as critical as in the case of beam-forming types, but element spacings are usually smaller. The merit of the high-mu gun depends on the extent of refinement of the structure; the smaller the wire size, the more turns per inch on the control-grid winding, and the smaller the cathode-to-grid spacing, the greater the merit. The price paid for the fine-grain structure is electrical fragility and more exacting mechanical tolerances. In general, the wire-to-wire clearance spacing should not exceed the grid-to-cathode clearance spacing.

The characteristics of the electron conduction system are shown by volt-ampere curves for the tube. These characteristics can be displayed in at least three different ways, as follows:

1. Constant-grid-voltage curves show the plate current obtained as a function of plate voltage for a family of different constant values of control-grid voltage. Fig. 30 shows this type of curve (called typical plate characteristics) for the 7214 beam power tube.

2. Constant-plate-current curves show the control-grid and plate voltages required to maintain a series of different constant values of plate current. Figs. 31 through 34 show typical constant-current curves for four different types of electron conduction systems.

3. Transfer characteristics curves show the relationship between plate current and control-grid voltage for various constant values of plate voltage.

The volt-ampere characteristics also usually include the control-grid and screen-grid current values as functions of the voltage values. In addition, for tetrode types several complete sets of characteristics are usually shown for different constant values of screen-grid voltage.

The space charge produced by the presence of electrons in the conduction space affects the voltage distribution substantially and must be considered in analytical computations concerning the design of the electron conduction system. Although many methods are available for computing the characteristics of space-current systems,[14, 34] they may sometimes be inadequate and empirical methods are also used.

Space-current systems may be studies by methods employing analogs. Resistance-network analogs can be constructed; the more complicated ones including space-current effects on voltage distribution require current injection. Ray tracing may be accomplished by means of "rubber dam" techniques. Recent advances in computers and their increasing availability may make it possible to predict the volt-ampere characteristics of any given system completely and accurately without actually building and testing the system. (See article "A New Approach to the Calculation of Electron Tube Characteristics" by O. H. Schade, Sr., in this book.)

Perhaps the most accurate method for designing the electron system is to construct a small sample or "unit" which can be easily assembled and disassembled, and in which spacings and shapes can be readily adjusted or changed. This unit can be evacuated in a demountable

vacuum system for measurement of its volt-ampere characteristics under actual applied-voltage conditions. This technique requires two special precautions: (1) A sufficiently good vacuum must be obtained so that the ionization of residual gas in the vacuum space does not neutralize the electron space-charge effects. (2) The processing must be of sufficient quality so that the electrode surface conditions are representative of those in the finished tube and the cathode is sufficiently activated to supply the required electron emission current without departing from the 3/2-power or space-current-limited emission value.

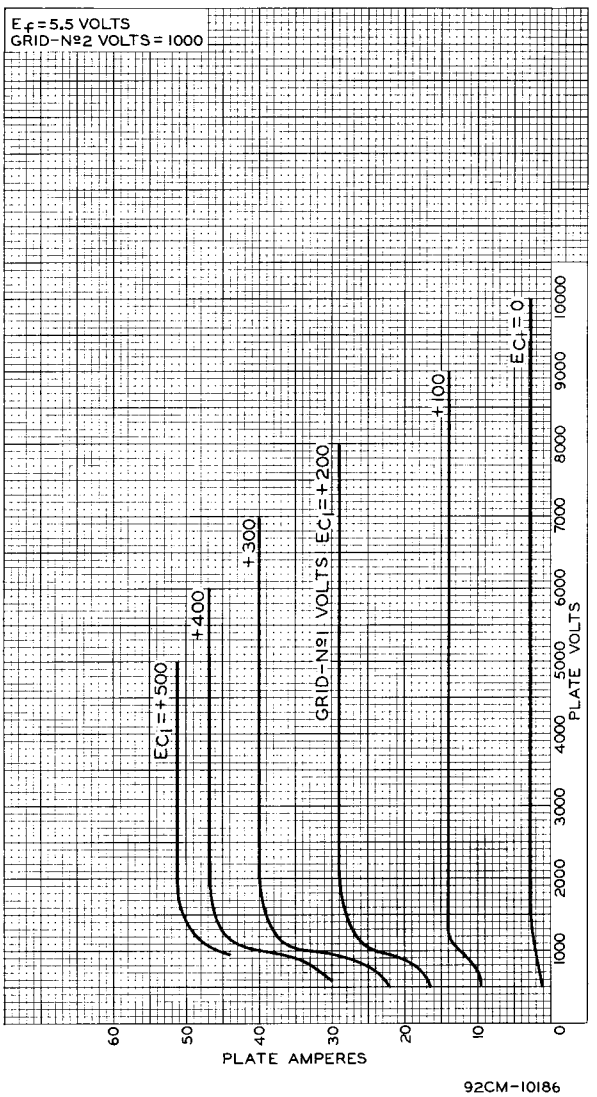


Figure 30. Typical Plate Characteristics for 7214 Beam Power Tube

In the design of electron conduction systems for control service, linearity of the plate characteristics is important because the output is often required to be a faithful reproduction of the input signal and to be free from distortion. In switch-tube service, the voltage drop across the tube should be a minimum during the current-conduction period to improve efficiency and to minimize the required dc plate-supply voltage. Low current absorption by the control grid (and screen grid) is also desirable. The characteristics and grid-dissipation capabilities should be such that the plate current is insensitive to minor changes in grid drive at saturation.

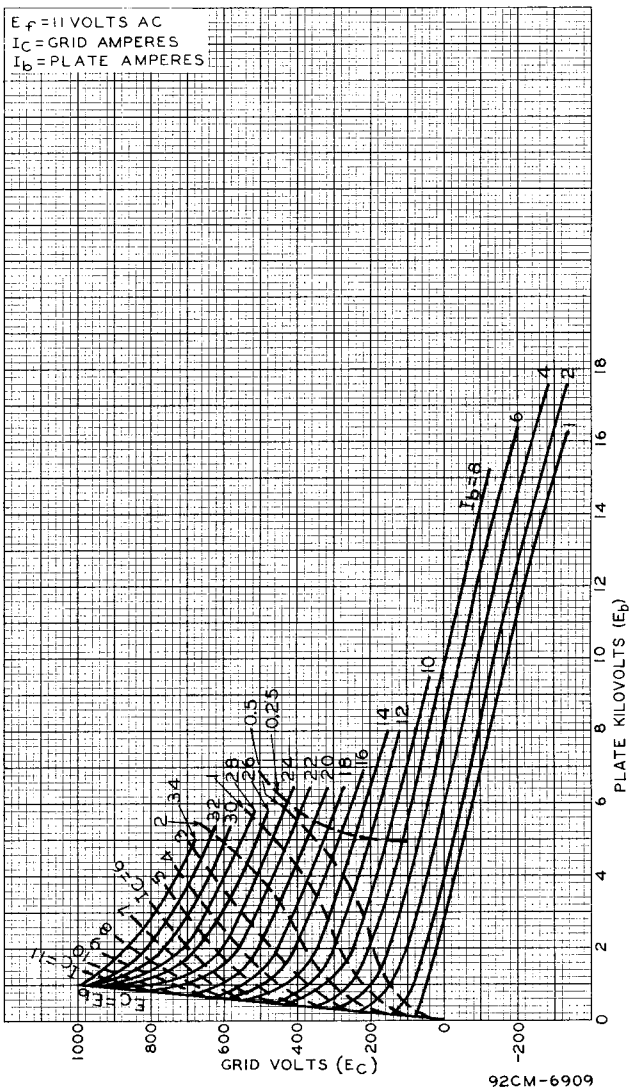


Figure 31. Constant-Current Curves for 5671 Power Triode

Because tubes in control service usually operate at high voltages and are expected to perform their control function under severe conditions without faulting, the electron conduction system should be able to withstand momentary overloading and should be rugged and not easily damaged by internal faulting.

Circuit Connections

Within the electron interaction region, the tube electrodes are used primarily for controlling electrons. Beginning at the edge of the electron beam where the interaction region ends, however, the electrodes serve a pure circuit or passive function electrically. Thus, the manner in which the electrodes are connected to the external circuit is extremely important in determining tube operation.

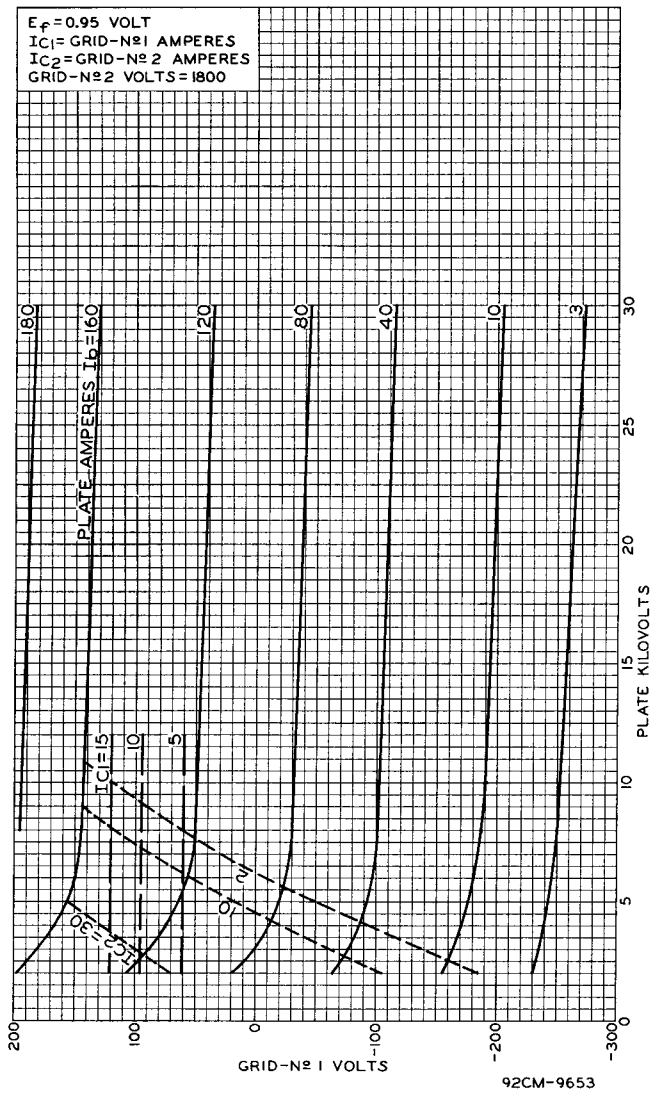


Figure 32. Constant-Current Curves for 6952 Beam Power Tube

There are many ways in which the elements of the electron conduction system can be connected to the external circuit. The simplest is the mere connection of wires, straps, or buss bars to the external tube terminals for completing the circuit paths; this method is used for very-low-frequency applications or for control service. In rf service, the tube terminals are connected to the lumped-constant elements of the rf tank circuits. Distributed-constant or transmission-line circuits may use parallel-line or strip transmission lines, or coaxial transmission lines (sometimes regarded as a cavity). In cavity resonators, the tube-circuit combination truly becomes a unit, i. e., the external circuit elements are connected to the tube electrode-supporting structures and envelope to form a complete self-shielding rf resonant circuit.[35, 36, 37, 38]

In some tube designs, the resonant circuit is completed wholly within the vacuum enclosure. In these integrated devices, the built-in resonant circuits have output terminals to which a pure resistive load can be attached to remove the generated power, and rf input terminals which present almost a purely resistive load to the driving source. These integral-circuit tubes are fixed-frequency devices unless some mechanical means (such as a bellows) is provided for moving parts of the circuit within the vacuum envelope to change the resonant frequency. Because no boundary conditions are imposed by the placement of ceramic seals, however, the circuit components can be tailored to fit the electron interaction system to provide the ultimate electronic bandwidth. The electronic bandwidth (signal bandpass achieved without mechanical tuning) may be further enhanced by means of coupled-circuit techniques,[39, 40, 41, 42] by intentionally offsetting the center frequencies of the resonant circuits in the amplifier chain in a prescribed manner, or by resistive loading of the resonant circuits to widen their response characteristics in the frequency spectrum. Any number of different types of high-frequency tube-cavity combinations may be executed.[12, 24-26, 28, 43-45]

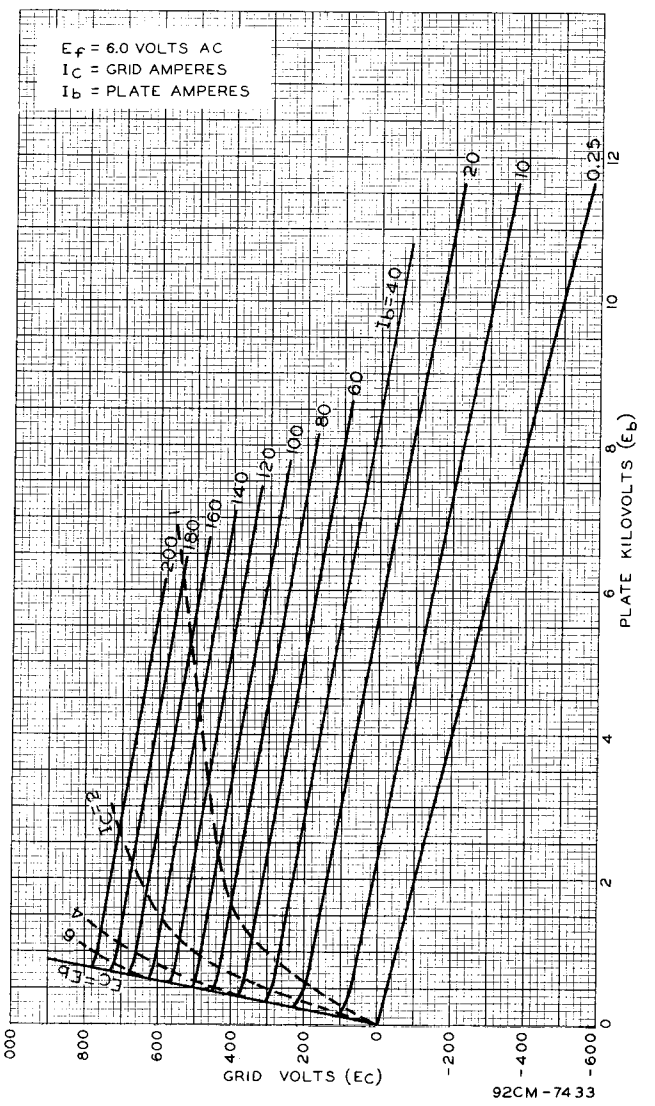


Figure 33. Constant-Current Curves for 5831 Beam Triode

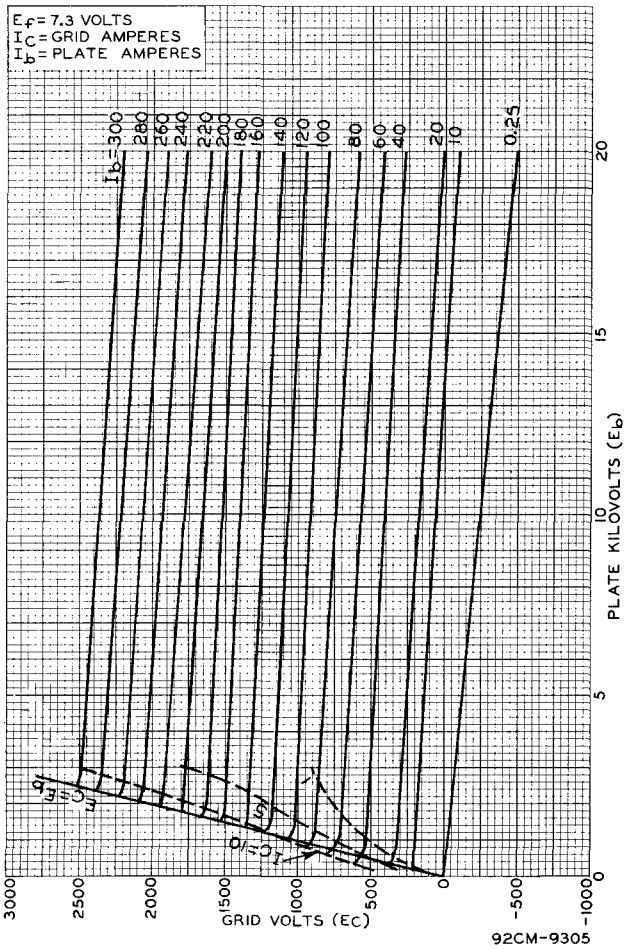


Figure 34. Constant-Current Curves for 6949 Shielded-Grid Beam Triode

The input circuit is almost always connected to the interaction space between the cathode and the control grid. Its function is to impress a signal voltage between these two electrodes of the proper magnitude to cause electrons to leave the cathode in sufficient quantities during all or a portion of the time that the instantaneous effective cathode voltage is positive. The pulses of electrons which originate at the cathode are then accelerated across the input space and pass as density-modulated pulses into the output space.

The input circuit, whether internal or external to the vacuum, should be designed for minimum circuit loss and thus maximum power gain. (The circuit loss, which is due to I^2R, should not be confused with electron current or interaction losses.) In other words, a given amount of driving power should produce a maximum grid-to-cathode signal voltage. This condition, which is usually obtained when the circuit is designed to store minimum energy, is also conducive to achieving maximum input-circuit bandwidth. If the bandwidth of the input circuit is too small, it may be increased by resistive loading; however, the over-all input-circuit efficiency will be reduced and the driving-power requirements will be increased.

Besides generating the voltage at the grid-cathode interaction space, the input circuit may also act as a transformer, i.e., it may transform the impedance at the interaction space into an appropriate value (usually resistive) for the driving source.

The output circuit is usually connected between the plate and the adjacent grid (in triodes, the control grid; in tetrodes, the screen grid). Its function is to utilize the bunches of electrons injected from the input region as efficiently as possible in establishing proper fields in the output space to convert dc power to high-frequency rf power. The output circuit is primarily a resonant or energy-storage system which establishes voltages on fields in the output space of the tube when electrons are injected in pulses. It also transforms the resistive impedance of the load and reflects a proper value at the output interaction space. Low passive or I^2R circuit losses result in greater circuit efficiency and delivery of maximum power to the useful load. Low energy storage in the circuit tends to increase the bandwidth.

In the conventional grid-driven tetrode, the circuit between cathode and screen grid is also important. For grid-drive operation, the screen grid and cathode must be at the same rf potential (i.e., no rf voltage between screen grid and cathode), and the same rf signal voltage must appear between control grid and screen grid as between control grid and cathode. Therefore, the impedance of the screen-grid-to-cathode circuit must be at least an order of magnitude lower than the impedance of the circuit or path between the screen grid and the control grid. Because the screen grid must be isolated from both the control grid and the cathode for dc voltages, the circuit must employ a bypass or blocking capacitor. This screen-grid-to-cathode capacitor must have low impedance at the required frequencies of operation. In addition, the remainder of the circuit path from screen grid to cathode emitting surface must be short and have very low inductive reactance.

As mentioned previously, coupling between the input and output circuits should be minimized. The use of a minimum number of openings in the metallic members separating the input and output circuits minimizes both inductive and capacitive feedback effects. Separate current-carrying circuits should also be provided for the input and output regions. These steps eliminate both feedback effects and possible unstable oscillation conditions at the fundamental operating frequency. They also reduce the possibility of feedback and oscillator conditions (and make it easier to eliminate them by suppression techniques) at harmonic frequencies, as well as at frequencies associated with other higher-order or lower-order modes and resonances which are inevitable in large circuit structures.

The discussion of input and output circuits, although generally applicable to all applications, has been directed primarily toward high-frequency rf service. It has also been concerned with relatively high-Q (class B and C) high-energy-storage circuits operating in the "standing-wave" mode. Another type of operation, analogous to the class A amplifier, is the "traveling-wave" mode, as used in the distributed amplifier. Such amp-

lifiers are usually low in efficiency, but they are outstanding for their wide bandpass capabilities and over-all power gain.[44]

A wealth of information is available in the literature on the design of both external circuits and those internal to the vacuum enclosure. The time has passed when the design of the electron conduction system can be separated from the design of the circuit. Increasing demands for higher power, higher frequency, wider bandwidth, higher over-all gain, smaller size, resistance to rigorous mechanical environments, and other special performance capabilities demand that the rf generator be designed as a complete integrated package in which the tube and the circuit function act as one operating unit.[12, 13, 24-26, 35, 46]

Dissipation Considerations

The dissipation capability of the grids themselves is probably the greatest single factor limiting the performance of grid-controlled tubes. The primary effects of inadequate grid-dissipation capability are evidenced by (a) thermal grid emission and resultant instability due to the flow of reverse grid currents in the circuits, and (b) overheating of grids and resultant mechanical failure, buckling, and even melting. In addition, there are secondary effects which create an unfavorable environment for the cathode; these secondary effects prevent realization of the inherent capabilities of the cathode for long life and high peak current.

In conventional power tubes, grids are heated by radiation[47, 48] from the cathode, by electron bombardment, and often by I^2R losses, but are cooled only by radiation or conduction through their support structures. Such tubes are almost always limited by their grid-dissipation capabilities.

Many methods have been tried to cure the "hot-grid" problem, including special high-thermal emissivity[47] grid surfaces, special processing techniques, and coatings which were intended to prevent grid emission even when grids become thermally hot. However, the only effective solution has been to keep the grid cold by conduction cooling. Recent tubes such as the 6448 tetrode[12] and the 2054 triode[10, 32, 49] make use of truly effective water-cooled grids; these tubes are not limited by control-grid or screen-grid dissipation capability.

As a result of the new conduction-cooled grids and the high-current capabilities of thoriated-tungsten and matrix-oxide cathodes, the limiting factor in many modern high-power tubes is plate dissipation. CW plate-dissipation capabilities have been increased by the use of high-purity water-system techniques[21] and improved plate designs which provide a better water-metal interface. In pulsed operation, however, severe thermal gradients are set up through the plate thickness from the bombarded surface to the water interface. These gradients cause severe mechanical stresses which reoccur at an acoustical rate (the repetition rate of the pulse) and may cause fatigue or eventual rupture of the plate material.

Copper is traditionally used as the plate material in power tubes because of its high thermal and electrical conductivity. However, it is a poor metal mechanically and has very low strength limit. Increasing the bombarded surface area of the plate reduces the unit dissipation level and has proved beneficial.[49] As the power level and energy concentration increase, however, it may be necessary to use more refractory metals such as molybdenum or tungsten for plate materials.

VACUUM ENVELOPES

In high-power, high-frequency tubes, the vacuum envelope can be a limiting factor on the maximum operating frequency and the maximum power capabilities. Particularly when the resonant circuit is partially within and partially external to the vacuum enclosure, the upper frequency may be limited by the accessibility of the external circuit. For example, if the active elements of the tube and the ceramic window are equivalent to 90 degrees of electrical length at a given frequency (the strap resonance), the tube cannot be operated in the $\lambda/4$ mode at or above this frequency. The $\lambda/2$ or $3\lambda/4$ mode can be used for operation at higher frequencies, but these operating modes tend to degrade the bandwidth performance of the tube.

When high power and high frequency are important, therefore, the tube designer should carefully consider the properties of the materials used in the ceramic-to-metal seals of the vacuum envelope, the bonding techniques used for the seals, and the seal configurations. The proper combination of these three factors makes it possible to minimize the critical electrical length.

Electrical Properties of the Ceramic

The important electrical properties of the ceramic materials[50] used in ceramic-to-metal seals include the following:

1. Dielectric strength is a measure of the voltage required to puncture the ceramic when it is placed between two electrodes. The "puncture" test differs from a surface failure in which the arc travels along the outer surface of the ceramic and does not enter the body of the material. Surface failures can be more troublesome because they are strongly influenced by the changing environment of the seal.

2. Dielectric constant is a measure of the increase in capacitance between two electrodes when the ceramic replaces a vacuum between them. Because the dielectric constant of ceramics ranges from about two to several hundred (high-alumina ceramics have a constant of about 9), their use produces a marked increase in capacitances. In addition, the electrical length in the direction of propagation is increased over its physical measurement by the square root of the dielectric constant. To conserve the electrical length, therefore, it is necessary to minimize the dimension in the direction of propagation and to choose as low a dielectric constant as possible.

3. Loss tangent is a measure of the heat generated within the ceramic as a result of dielectric losses. It is the ratio of the electrostatic losses to the elec-

trostatic energy stored per cycle, or, more exactly, the tangent of the angle between the real component of current through the ceramic and the imaginary component.

4. Loss factor is a measure of the total loss within the ceramic, and is equivalent to the reciprocal of the Q of a capacitor. It is the product of the dielectric constant and the loss tangent. The dielectric constant provides a measure of electrostatic energy stored, and the loss tangent gives the ratio of losses to that stored energy.

Mechanical Properties of Ceramics and Metals

Both the ceramic and the metal elements of the vacuum enclosure must possess certain mechanical properties to provide satisfactory seals. The following properties are the most important:

1. Expansion rate. The difference between the expansion rates of the metal and ceramic parts determines the rate of increase in stresses with temperature changes. If the two expansion rates were perfectly "matched" and the seal were in thermal equilibrium, a very thin ceramic could be used without fear of stress failure.

2. Mechanical strength. Obviously, the mechanical strength of both the metal and ceramic parts must be greater by some safety factor than the stresses induced by the extremes of the environmental temperatures. Because the compressive and tensile strengths are equal in metals, the elastic stress limit and the ultimate stress values must be considered. In the region of elasticity, stress increases linearly with strain, and thus with expansion differences. For this reason, soft annealed copper is often used as a strain-isolation member in seal designs. In ceramics, the ultimate stress and the elastic limit are essentially equal, but the tensile strength is very low as compared with compressive strength.

3. Moduli of elasticity. These moduli determine, to some extent, the degree of sharing of the strains, and therefore the stresses, between the metal and ceramic parts. This factor is important in a complete stress analysis of seal design, but is less important than the two previous factors.

4. Thermal conductivity. The thermal conductivity of both metal and ceramic elements must be considered to provide for the removal of all heat losses generated in the seal assembly. This factor is particularly important in the ceramic, where localized losses can produce sufficient thermal gradients to produce excessive tensile stresses.

Bonding Techniques

A detailed description of sealing and bonding techniques is given elsewhere in this book (see article "Ceramics and Ceramic-to-Metal Seals" by M. Berg); however, a brief summary and comparison are included here with special reference to the techniques used in large power tubes.

Because ceramic is an oxide, it can not be "wet" by molten metal, and thus can not be bonded by conventional brazing techniques. The techniques used most frequently for ceramic-to-metal seals in power tubes are reactive-metal brazing, diffusion of refractory metals, and mechanical or compression assembly.[51, 52, 53]

In reactive-metal brazing, a highly reactive metal such as titanium is used to reduce the surface layer of the oxide (ceramic) so that wetting may occur. The titanium or titanium hydride is generally mixed with the brazing alloy, and the assembly is mounted in a jig for conventional brazing. However, the brazing must be done in vacuum or in an inert-gas atmosphere to prevent oxidation of the titanium before it can reduce the ceramic.

In the diffusion method, a refractory metal such as molybdenum or tungsten is diffused onto the surface to be brazed. Usually, the ceramic is painted with an ink containing the fine metal powder, and then fired in wet hydrogen at temperatures of 1400 to 1500 C until the diffusion is sufficiently deep. The surface is nickel-plated to improve its wettability, and the metallized ceramic is then brazed in a conventional manner. Typical applications of this metallizing technique are shown in Fig. 35.

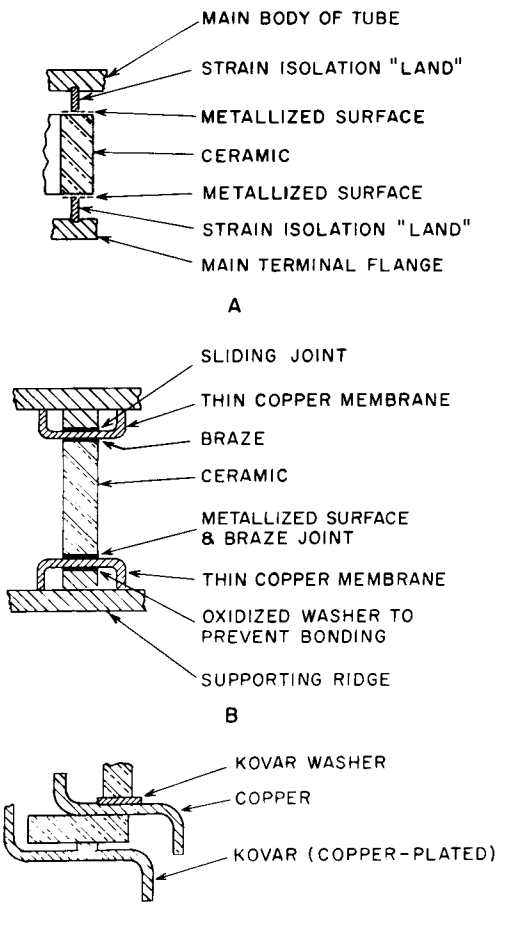


Figure 35. Examples of Metallized and Brazed Ceramic-to-Metal Seals

Both the reactive-metal and the diffusion techniques depend on molecular adhesiveness or chemical interaction for the ceramic-to-metal bond. The compression seal,[18, 54] however, is simply a press fit between metal and ceramic. In this seal, deformation or strains created during assembly cause radial compressive forces on the interface which produce the vacuum-tight seal. Elastic stresses then maintain these forces during processing and tube operation. Typical compression-seal configurations are shown in Fig. 36. In general, this technique is used when a rugged seal is required to handle very high average powers.

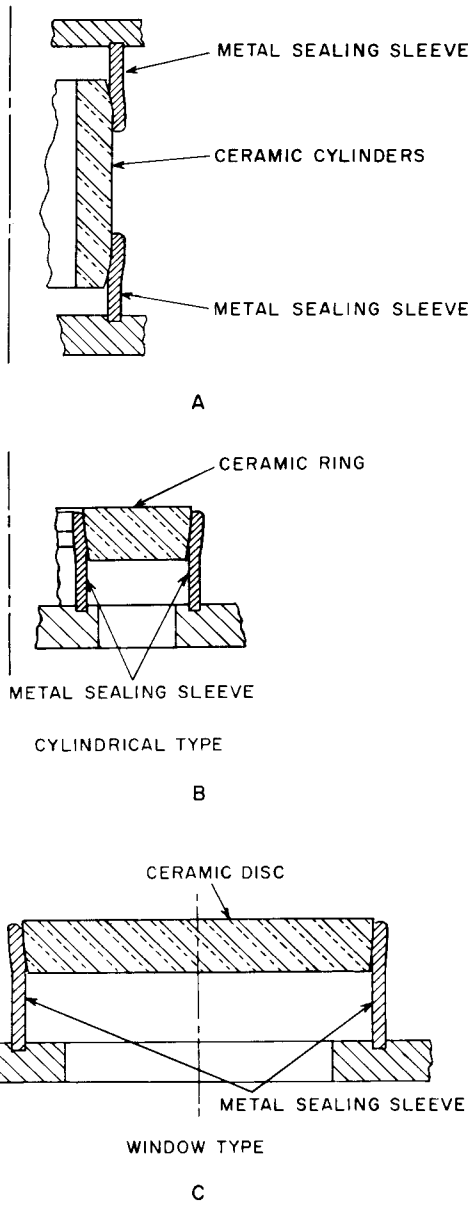


Figure 36. Typical Compression-Seal Configurations

Strain Isolation

In both the metallized seal and the compression seal, some length of metal is used as a strain isolator to compensate for the different expansion rates of the ceramic and metal parts. The compression seal uses a hardened-steel sealing sleeve of proper length and thickness to isolate and yet maintain sealing forces throughout the required temperature range. The steel has poor thermal conductivity, but high rigidity. The metallized seal uses either soft copper or kovar as an isolator. Copper has good thermal conductivity, but that of kovar is very poor.

The isolator shown in Fig. 35A is simply a thin sleeve sufficiently long to relieve the differential strain in its plastic deformation. The type shown in Fig. 35B is similar, except that the copper foil used is too thin to support itself; consequently, a supporting ridge is placed beneath it to provide axial support. This seal provides the simplest rf path (least field disturbance), but has little rigidity in the plane perpendicular to the axis. When copper is used in either of the isolators shown in Figs. 35A and B, it is made sufficiently thin so that it can not create sufficient stress to crack the ceramic.

Fig. 35C shows two isolators used in small power tubes; the upper seal uses a kovar washer to restrain the forces of the copper; the lower seal uses matched (copper-plated) kovar entirely. The small emboss in the lower seal is used to confine the width of the braze to a narrow region.

Surface Losses

Because the compression seal is purely mechanical, there is no adulteration of the surfaces and the surface resistivity is the same as that of the copper or silver plating used. In the metallized seal, however, surface resistivity is increased as a result of the interaction of the oxide and the metal. The extent of this increase varies with the type of metallizing and the technique used. A report from the University of Illinois[55] indicates that molybdenum-manganese has the same resistivity as nichrome. Work in progress at Lancaster indicates that this resistivity can be reduced by proper metallizing techniques. The main considerations are the magnitude of the loss and the possibility of sufficient cooling if the loss does not seriously affect the circuit efficiency. This consideration becomes more important at higher frequencies, where circuit efficiency tends to decrease and accessibility for cooling becomes more difficult.

Assembly Process

The compression seal is made in three steps:[51, 52, 53] (1) the ceramic is ground to close tolerance (if small parts are pressed isostatically, grinding may not be necessary). (2) The sealing sleeve is polished. (3) The ceramic and sleeve are pressed together in a hydraulic press.

For the metallized seal, grinding of the ceramic may not be required if tolerances can be held sufficiently close. This seal is then made as follows: (1) The surfaces of the ceramic are painted or printed with metallic ink and fired in wet hydrogen at 1400 C. (2) The metallized surface is nickel-plated. (3) The parts are jigged and brazed in a hydrogen furnace.

Voltage Standoff

Both the metallized and the compression seal must be shielded in some way from excessive voltage gradients if maximum tube capabilities are to be achieved. Seals have a tendency to initiate corona, which then may cause flashover or, in the case of the metallized seal, vaporize the silver from the feather edge of the metallizing. This vaporized silver may then be deposited on the ceramic and eventually cause breakdown.

Fig. 37 shows three types of shields which can be used for corona suppression.

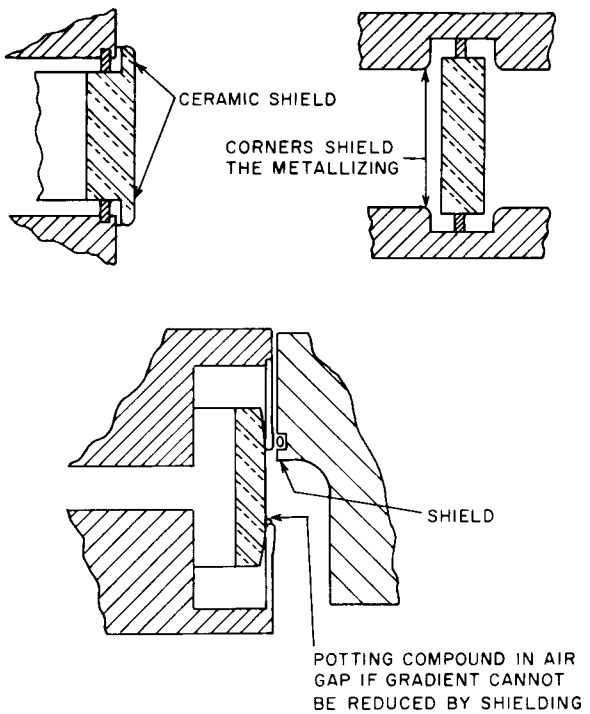

CERAMIC SHIELD
CORNERS SHIELD THE METALLIZING
SHIELD
POTTING COMPOUND IN AIR GAP IF GRADIENT CANNOT BE REDUCED BY SHIELDING

Figure 37. Shields Used for Corona Suppression

Materials

The high-alumina ceramics are very well suited for insulators and rf windows in vacuum enclosures because of their high mechanical strength and low rf loss factor. Beryllium oxide, when it becomes more readily available, will be even better because of its high thermal conductivity. Pyroceram, the crystalline glass, offers the advantage of a close expansion match to metals, as well as the possibility of more economical forming techniques.

MECHANICAL ASPECTS OF TUBE DESIGN

Because the power tube is an electrical device, its theoretical design is based on electrical requirements which become the governing parameters for the design. However, the degree to which the calculated electrical performance can be achieved is directly related to the solution of mechanical problems. For example, the use of closer and more uniform control-grid-to-cathode spacings to obtain tubes of higher merit depends entirely on the success of mechanical design and execution. Good tube design must be based on an underlying mechanical philosophy of construction and assembly rooted in sound physical and mechanical premises.

There are two facets to the mechanical design requirements. First, the tube must meet the basic requirements of normal operation and environment, including normal operating temperatures, thermal stresses, and thermal cycling. If required, it must withstand environmental conditions of shock, vibration, and extreme temperatures without detrimental irreversible change which might impair its useful life.

Second, the tube design must be compatible with available construction techniques. The parts must be designed so that they are capable of being fabricated and assembled, and also of withstanding the rigors of high-temperature brazing or other joining methods such as welding. The finished tubes must be able to withstand the rigors of final processing, in which they are submitted to temperatures, dissipation conditions, and thermal stresses beyond those of normal tube operation.

Besides the specification of shapes and dimensions of component parts and assemblies of the tube, mechanical design usually entails the selection of material and joining methods to be used. These two items represent a vast area of tube technology which is particularly relevant to power-tube design. The selection of materials for use in power tubes requires consideration of most of their physical properties. When dissimilar pieces are to be joined, it is necessary that a certain pattern of compatibility exist among the components of the assembly to effect a satisfactory union. In a simple vacuum-tight braze joint, for example, if the difference in thermal expansion between the two parts is so great that one of the parts is stressed beyond its ultimate strength, the part will fracture and an unsatisfactory assembly will result.

When alloys are used, an understanding of the metallurgy of the metals is essential.[56] Alloys are useful in many instances because of the strength and hardness which can be achieved as a result of special heat-treating schedules. The designer should know what happens to the metals when they are subjected to a given set of time-temperature treatments.[57, 58]

It is important to realize that when two metals are in contact and one or both of them go into solution, new alloys and mixtures of metals result. In many cases, these new alloys have inferior mechanical and electrical properties. Again, compatibility of the metals and the joining technique is important.[51, 56, 59]

Electro-plating of metal parts may be used to impart special surface properties to structural parts so that brazing can be carried out, or to provide a high-electrical-conductivity surface.

The types of joining methods commonly employed in power tubes are as follows:

1. Welding, which is used for similar metals and, with caution, for dissimilar metals.
2. Brazing, in which two parts are joined together

by a third metal or alloy having a lower melting point than either of the two parts being joined.

3. Diffusion, in which two metallurgically compatible parts are held together under pressure and heated for an extended period of time to a temperature substantially below the melting point of either part.

4. Mechanical or force-fit joints, in which residual stresses within the parts maintain the joint.

The tube designer should have a knowledge of the physical, as well as the electrical, properties of the range of materials used in power tubes. Some of the important physical properties include:

composition[51, 57, 60-63]
thermal conductivity[57, 60, 62-67]
electrical conductivity[57, 60, 63, 66-70]
density[57, 60, 62, 67]
magnetic properties (permeability)[57, 60, 63, 66, 67, 71]
thermal expansion as a function of temperature[57, 60, 62-64, 66, 67, 72-74]
vapor pressure as a function of temperature[57, 60, 67, 75, 76]
melting point[57, 60, 62-64, 66, 67, 73]
hardness (for specific thermal and/or mechanical treatment)[57, 58, 62, 64, 67, 73]
ability to withstand voltage[50, 77]
critical metallurgical temperatures where phase changes occur (for alloys)[51, 57, 60, 62, 65]
strength:[57, 58, 60, 62-64, 66, 67]
- elastic limit (at low and high temperatures)
- yield point
- ultimate strength or rupture point

modulus of elasticity

The following optical properties are also useful:

spectral emissivity (or curves showing true temperature as a function of optical brightness temperature[48, 58, 60, 67, 78]
total emissivity

The properties listed above apply particularly to metals, although some are also applicable to dielectrics (insulators). Two additional properties of dielectrics are also important: [50, 60, 63, 70] dielectric constant as a function of temperature and frequency loss factor as a function of temperature and frequency.

The designer should, of course, also have a working knowledge of the various materials used in different tube components. Some of the more commonly used materials are listed below:

For filaments, heaters, and cathodes
- pure tungsten[79, 80]
- thoriated tungsten
- tantalum[81, 82]
- nickel[83, 84]
- Hastalloy B (a precipitation-hardening nickel alloy)[57]

For grids
- oxygen-free high-conductivity (OFHC) copper[57, 60, 62, 63, 67, 73, 85]
- precipitation-hardening copper alloys:[57, 86]
  - chrome copper
  - phosmic bronze (#58 alloy)[87]
  - pure tungsten[63, 79, 80]
  - platinum[88]
  - molybdenum[57, 62, 63, 67]

For plates (anodes)
- OFHC copper[57, 60, 62, 63, 67, 73, 85, 86]

For vacuum envelopes
- aluminum oxide[50]
- beryllium oxide[50]
- glass[50]
- OFHC copper[85, 86]
- various tool steels[58]
- kovar[63] (Fermico and cobalt nickel-iron alloy having poor physical and electrical properties but having a controlled thermal expansion rate which matches that of certain hard glasses)
- Inconel and Inconel X (precipitation-hardening nickel alloys)[89]
- stainless steels[57, 72, 90]

For miscellaneous parts, supporting structures, brazing alloys, etc.
- silver[57, 67]
- gold[57, 67]
- high-speed tool steels (18-4-1)[58]
- silver-copper brazing alloy[51, 63]
- gold-copper brazing alloy[51, 63]
- sapphire[50]
- quartz[50]
- mica[91]
- beryllium copper[57, 86]
- precipitation-hardening stainless steels[57, 89, 90]
- zirconium[92]

EXHAUST AND PROCESSING

During the exhaust and processing cycle, the tube is transformed from its unevacuated yet mechanically complete form to a highly evacuated, metallurgically and chemically stable, sealed-off state ready for final electrical tests.

The major problem in achieving suitable high-vacuum conditions for a power tube (10^{-6} to 10^{-8} millimeters of mercury) is related to the condition of the internal surfaces exposed to the vacuum. (A special case is the oxide-cathode tube, where chemical breakdown of the cathode surface liberates great quantities of gas.) It is not particularly difficult to achieve a vacuum of 10^{-8} millimeters of mercury in an enclosed chamber; the problem is to maintain the vacuum under tube operating conditions, when heated surfaces, high-voltage gradients, and electron bombardment drive off adsorbed gas particles from "unclean" surfaces exposed to the vacuum.

The principal objective of the processing cycle, therefore, is to outgas the internal surfaces of the tube so that subsequent operation will not impair the quality of the vacuum. This outgassing is accomplished by heating the parts, by electron bombardment, and by the application of high voltages. Other processes, such as pro-

ducing a controlled glow discharge within the tube, can also produce beneficial "cleaning" results.

There is no substitute, however, for minimizing all sources of contamination of tube parts in the first place. Throughout the assembly process, extreme care should be taken to keep the parts clean and free from all foreign material. Cleaning methods and cleanliness tests are discussed in the literature.[93] Generally, it is undesirable to use cleaning methods involving chemical reactions on final and partial tube assemblies because the byproducts of such reactions can be sources of contamination.

Bakeout

The first step in the exhaust cycle consists of a high-temperature bakeout. The tube is placed in an oven and heated slowly to the bakeout temperature, which is usually 350 to 500 C. This temperature is maintained for a period ranging from 12 to 24 hours while the tube is continuously pumped out by an appropriate high-vacuum pumping system. (See article "High-Vacuum Technology and Equipment" by W. G. Henderson and A. L. Lucarelli in this book.)

The purpose of the bakeout cycle is to drive off occluded and absorbed gases from the internal metal and insulator surfaces of the tube. This process is accelerated by heating the parts, and is more thorough and more complete for higher temperatures and longer bakeout times. As a result, tubes are usually baked out at the highest temperatures the materials and structure can tolerate, or at the limiting capability of the bakeout equipment. Many large tubes are maintained in a secondary vacuum chamber during the bakeout cycle to prevent excessive oxidation of the external surfaces at the bakeout temperature and to eliminate a subsequent cleaning process.

Heating the Electrodes

After the bakeout cycle is complete, the tube electrodes are heated for extended periods to temperatures above their normal operating values. Again, it is generally true that the higher the processing temperature and the longer the time at the high temperature, the more thorough is the processing and the better the ultimate vacuum in the tube. However, if the temperatures are too high or the time too long, irreversible mechanical, metallurgical, and chemical changes may occur which will damage the tube or impair its characteristics. The secret of successful exhaust processing is to find the combination of temperature and time which is most beneficial in outgassing the internal tube parts without damaging them.

In the processing of thoriated-tungsten tube types, the filaments are outgassed first, with the other elements of the tube heated only to the extent of the power radiated from the filament. The plate, in particular, is not cooled during this step, but its temperature is allowed to rise. After the filament is outgassed, voltages are applied to the other electrodes and they are further heated by electron current drawn to them from the hot filament.

In matrix-cathode tubes, great quantities of gas are liberated from the cathode during processing. When the cathode is heated, a chemical change takes place in the coating during which the carbonates of rare earths are chemically converted to the oxides of the metals. Decomposition products released by the cathode during this "activation" process have a tendency to deposit on cooler surfaces of the tube. These decomposition materials are undesirable on other active electrode surfaces, particularly the plate, which is later bombarded by electrons. In these tubes, therefore, the plate is heated to a high temperature (700 to 800 C) first, and is maintained at this temperature while the cathode is heated and activated. Decomposition products given off by the cathode then can not stick to the plate surface, but are pumped out of the tube. Matrix-cathode tubes usually require a series of electrical processing steps to activate the semiconductor emission material properly.

Except for the cases noted above, it is usually desirable to heat all the electrodes simultaneously to avoid chasing the decomposition products from one electrode to another. The last processing step prior to pinch-off (removal from the exhaust system) should be a simultaneous heating of all electrodes; in this step, the tube should be as hot as possible, even to the point of having the filament lighted when the final pinch-off is made.

High-Voltage Breakdown

In the exhaust processing of high-voltage power tubes, high voltage is normally applied between the plate and other electrodes while the tube is "cold" (no filament or heater voltage applied). During this processing, minute particles or sharp points within the tube are subjected to intense voltage gradients. The resulting electron field emission and moving charged particles cause a flow of current (called dark current) across the vacuum space.[77, 94, 95] As the voltage is increased, the intensity of this dark current in a local area may increase to the point where a breakdown occurs and intensely localized energy is released. Such breakdowns change the local surface conditions of the electrodes, reducing gradients, rounding off sharp points, eliminating loosely adhering particles, and thereby improving the voltage hold-off capabilities of the tube.

The voltage value at which breakdown occurs, as well as the frequency of breakdown, also depends on the rate at which the voltage source (power supply) can supply energy to the fault condition. Many occurrences of dark current do not cause a breakdown or "fault" because the power source can not supply energy fast enough to maintain the current. Power supplies with low internal impedance and high-current capabilities cause breakdown at lower voltages. The capabilities of the power source may be increased by attaching a good storage capacitor across the output terminals near the tube being processed.

Tubes intended for switching service are subjected to the most rigorous voltage-standoff conditions because they are connected across the terminals of a large storage-capacitor bank having extremely high current and energy capabilities. In such an arrange-

ment, the slightest internal tube activity can cause a flash arc or internal fault. Consequently, these tubes require the most vigorous processing with respect to high voltage and the energy capability of the supply.

Because the voltages used for high-voltage treatment (50 kilovolts or more) can produce substantial dangerous X-radiation, particularly when substantial dark currents are flowing, proper safety measures must be taken to prevent exposure. These measures may include the use of lead shielding, and the regular checking of film badges worn by personnel subject to exposure.[96]

Activation and Aging

After the tube is removed from the exhaust equipment, it is operated under dynamic conditions, first at low power-voltage levels and then at increasingly higher power levels until the final aging conditions are reached. Final operating conditions are usually at or above the maximum published ratings for the tube.

The rapidity with which the power level is increased depends upon the response of the tube. Gas current should be monitored at frequent intervals to avoid tube operation under extremely poor vacuum conditions. Many flash arcs or internal tube faults occur during this processing period; if they cause the cathode emission to decrease, the tube may be operated with only filament power applied but at greater-than-normal filament temperatures to reactivate the cathode coating.

This aging process is continued until the tube has reached a stable condition as far as emission capability, gas evolution, and flash arcs are concerned, and can be operated at the full level of the high-power test conditions without a tube fault for one hour or more.

Fault Protection

Microsecond fault-protection techniques must be used during the aging process, when so many tube faults normally occur, to protect the tube structure from permanent damage.[97] The rectifiers employed in super-power testing equipment are powerful devices with extremely high short-circuit current and power-delivery capability. If these energies are concentrated in a localized spot within the tube when an arc occurs, severe damage to the structure may result unless the power source can be disconnected within a few microseconds.

Mechanical circuit breakers and grid-controlled rectifiers require periods of the order of several milliseconds to cut off power. It is feasible, however, to divert the fault from the tube within a few microseconds by firing an ignitron connected across the power-supply bus. The ignitron is normally held in a non-conducting condition. When the normal power-interrupting mechanism, either a circuit breaker or grid blocking of the rectifier tubes, is caused to function, it interrupts the flow of current in the ignitron (sometimes called "crowbar" or "dump" tube).

Fault protection is mandatory for all tube processing in which the energy in the supply can damage the internal tube structure in case of a fault. In addition, such circuits are always recommended for use in high-power equipment.

Getters and Getter Ion Pumps

Because the activation and aging processes are carried out after the tube is removed from the exhaust equipment, gas generated during these processes is merely redistributed among the various surfaces within the tube and not pumped out. As a partial remedy for this problem, most large power tubes employ a getter, a chemical pump formed by evaporating an active layer of barium compounds on a glass surface within the tube. Most gases within the tube react chemically with this getter surface and adhere to it. Another gettering method is to attach a piece of zirconium metal to an appropriate heated internal tube part, such as a radiation-cooled grid member.

In many tubes, a preprocessed type of getter is used which consists of a conveniently shaped glass container loaded with the getter material. The getter is "flashed" and the getter material evaporated on the inside surface of the glass container by conventional rf induction-heating methods. The flashing step is best carried out with an atmosphere of argon gas in the container; the active getter surface produced in argon is much lighter and "fluffier" than the smooth surface produced in a vacuum, and thus provides a greater total surface area for gettering action and a greater total pumping capacity.

More effective continuous pumping action may be provided by the use of a getter ion pump.[98, 99] These pumps are small in size, have no mechanical moving parts, and require a power source of only a few thousand volts dc and microamperes of current. They permit operation of the tube under full-power dynamic conditions with the pump still attached to the tube to remove gas generated during processing. This technique is particularly applicable to high frequency tube types in which the cavity circuits are built as integral parts of the tube within the vacuum enclosure. Conventional processing techniques are difficult to apply to such tubes, and often heating and bombarding certain parts of the tube can be accomplished only by high-frequency dynamic operation. The principal advantage of the ion pump in such cases is its small size and portability.

TEST AND EVALUATION

Throughout the development program for a new tube type, comprehensive tests are performed to establish the validity of the design. Enough test samples must be used to establish tube operating ratings and conditions for the intended type of service, and a tube-testing sequence for use on production units.

Two basic preliminary tests are performed on all tubes. First, the "cold" tube is tested for any internal short circuits between electrodes which are designed to be open circuited. In this test, sufficiently high voltages are applied between electrodes from a high-impedance power supply to establish that no permanent

shorts exist and that a good vacuum warrants the application of filament power. The purpose of the second test is to determine the gas pressure (degree of vacuum) within the tube. In this test, the tube is operated as an ion gauge, i. e., electrons are drawn to a positive electrode and ionized gas particles are collected at a negative electrode. When a getter ion pump is used, this test may be eliminated and the base pressure measured with the pump.

Over-all Evaluation

In the over-all evaluation of a new tube type, a series of basic independent tests are performed to establish compliance with individual performance requirements. These tests involve measurement of the parameters upon which the original design was predicated, such as:

1. cathode-emission capability,
2. voltage hold-off capability between electrodes,[77]
3. ability of various coolant passages or air ducts to pass the required fluid flow at the anticipated pressure drop,
4. volt-ampere characteristics (ability to pass sufficient electron current to the plate at the anticipated voltages),
5. dissipation capabilities of the various electrodes.

Additional tests performed on high-frequency tubes to measure their properties as circuit elements include the following:

1. tuning characteristics showing the dimensions of a specific cavity shape as a function of the resonant frequency of the tube and cavity (for a fixed-tuned device, the fixed frequency is measured);
2. unwanted modes or resonant conditions within the desired operational frequency bands caused by structural elements of the tube;
3. other properties such as circuit loss, loading characteristics, etc.[35, 36, 38, 46]

Life tests are also performed on the critical elements of the tube, particularly the filament and/or cathode and its supporting system. These tests are conducted for extended periods of time, both with the filament operating under steady-state conditions to maintain normal operating temperatures, and with the filament voltage cycled at a rate which allows the filament to attain thermal equilibrium in both the on and off conditions. The purpose of these life tests is to demonstrate that the emission capabilities are maintained at an adequate value, that the filament volt-ampere characteristic does not change substantially, and that the mechanical properties (structural integrity) of the device do not deteriorate throughout the expected service life. Special life tests may also be required for components such as the envelope of the anode if they are novel in design or will encounter operating conditions beyond the current state of the art.

Comprehensive over-all performance tests are also made under the conditions of and in an environment closely simulating that of the intended application. The tests of primary importance are the demonstration of the required power output at the required operating frequencies. Tests for other performance criteria, such as bandwidth, linearity, noise generation, and phase stability, may also be important, depending on the application. The final proof of performance consists of a life test of sufficient duration conducted under the conditions and environment of the intended application.

Diagnostic Tests

Another important type of testing is used when a tube type under development is not performing as anticipated and the reason for the malfunction is not evident. In such diagnostic tests. a cause is hypothesized and a testing sequence is carried out to determine the validity of the hypothesis. If the hypothesis is correct, measures can be taken to correct the malfunction. If it is wrong, a new one is proposed and the process is repeated.

In many cases, it may be impossible to obtain a positive indication concerning the validity of the original hypothesis. However, the test sequence may provide new information which will be useful in proposing a new hypothesis, or analysis of the test data may reveal a new and useful approach to the problem. It is important, therefore, that each test be carried through to a logical conclusion. It is often useful to plot the results obtained in curve form and study them carefully.

Mechanical Tests

Certain mechanical tests are also important in the evaluation of a new device. For example, the physical dimensions of the final design must be established and defined within tolerances that insure mechanical compatibility with the intended equipment. Gauging procedures which simulate the mechanical environment are often the best means for defining and checking the mechanical dimensions. The device may also be subjected to specified conditions of shock and vibration to simulate possible environmental conditions of the application.

An important point, of course, is that the tube be delivered to its user in operable condition. Therefore, the tube and its shipping carton must be designed to withstand transportation hazards. A program of packaging drop tests and actual shipping tests of the tube in its container is an essential part of the tube design program.

Circuit Development

It is often necessary for the tube design team to develop specialized circuits and test equipment for use in the development and final evaluation of the power tube being designed.[1, 15, 42] Of particular importance is the development of rf cavity circuits for use with tubes of the external cavity type. Because cavity design and tube design draw on many of the same disciplines, it is desirable to have the same engineering team develop the tube-cavity combination. Some of the principles applicable to this area include:

1. choice of configurations to minimize voltage gradients,
2. elimination of solid insulating materials wherever

possible, or judicious shaping and location of such materials to minimize high-voltage gradients,

3. elimination of sliding contact fingers wherever possible, or selection of designs which minimize current flow through the contacts. 25, 26, 35, 36, 37, 43, 46

REFERENCES

1. Yingst, T. E., "Circuits to Control and Protect High-Power Modulator Tubes," Electronics, January 12, 1962
2. Terman, F. E., Electronic and Radio Engineering, Fourth Edition, McGraw-Hill Book Co., New York, p. 525
3. Brown, H. H., C. N. Hoyler, and R. A. Bierwrith, Theory and Application of Radio-Frequency Heating, D. Van Nostrand Co.
4. Weldon, J. O., "A 600-Kilowatt High-Frequency Amplifier," Trans. IRE, PGCS-5, No. 1, March 1957
5. Winningstad, C. N., "Generating RF Energy for 6-BEV Bevatron," Electronics, Vol. 29, p. 164, February 1955
6. Hobart, T. D., "Navy VLF Transmitter Will Radiate 1000 Kw," Electronics, Vol. 25, p. 98, December 1952
7. Wagener, W. G., "Simplified Methods for Computing Performance of Transmitting Tubes," Proc. IRE, Vol. 25, No. 1, Part I, January 1937
8. Ayer, R. B., "Use of Thoriated-Tungsten Filaments in High-Power Transmitting Tubes," Proc. IRE, Vol. 40. p. 591, May 1952
9. Wolker, H. S., and W. H. Aldous, "High-Power Transmitting Valves with Thoriated Filaments for Use in Broadcasting," Proc. IEE, Vol. 107, No. 32, Part B, March 1960
10. Hoover, M. V., "Advances in the Techniques and Applications of Very-High-Power Grid Controlled Tubes," Proc. International Convention on Microwave Valves, May 1958
11. Spitzer, E. E., "Grounded-Grid Power Amplifiers," Electronics, Vol. 19, p. 138, April 1946
12. Bennett, W. P., "A Beam Power Tube for Ultra-High-Frequency Service," RCA Review, Vol. 16, No. 3, p. 321, September 1955
13. Salisbury, W. W., "The Resnatron," Electronics, Vol. 19, No. 2, p. 92, February 1946
14. Schade, O. H., Sr., "Beam Power Tubes," Proc. IRE, Vol. 26, No. 2, p. 137, February 1938
15. Shrader, M. B., "10-Kw Air-Cooled Tetrode for VHF Television Service," Tele-Tech, December 1953
16. Fisk, J. B., H. D. Hagstrum, and P. L. Hartman, "The Magnetron as a Generator of Centimeter Waves," Bell Syst. Tech. Jour., Vol. 25, No. 2, p. 342, April 1946
17. Chodorow, M., E. Ginzton, I. Neilsen, and S. Sonkin, "Design and Performance of a High-Power Pulsed Klystron," Proc. IRE, Vol. 41, p. 1584, November 1953
18. Grimm, A. C., "The Development of a High-Power Klystron Amplifier Tube," Final Technical Report, Contract AT(04-3)-363, Sub-Contract S-126, April 1962
19. Chodorow, M., and E. J. Nolos, "The Design of High-Power Traveling Wave Tubes," Proc. IRE, Vol. 44, p. 649, May 1956
20. Pierce, J. R., Traveling Wave Tubes, D. Van Nostrand Co.
21. Martin, I. E., "Evaluation of Coolants for High-Power Transmitting Tubes," Proc. National Conference on Tube Techniques, September 1958
22. Code XI
23. Beurtheret, C., "The Technique of Vaportions," Thomson-Houston Review Electronique CFTH, December 1956
24. Code X2
25. Koros, L. L., "A Novel Ultra-High-Power Amplifier System," RCA Review, Vol. 16, No. 2, June 1955
26. Parker, W. N., "The Continuation of the Development of a High-Power Broadband, Integral-Cavity Coaxitron Amplifier," Final Technical Report, Contract AF30(602)-2365, RADC-TDR-61-315, November 1961
27. Bennett, W. P., E. A. Eshbach, C. E. Haller, and W. R. Keye, "A New 100-Watt Triode for 1000 Mc," Proc. IRE, Vol. 36, p. 1296, 1948
28. Peterson, F. W., "A New Design Approach for a Compact Kilowatt UHF Tetrode," 1958-IRE Wescon Convention Record
29. Weber, E., "Linear Transient Analysis," Vol. 1, p. 244, John Wiley, New York
30. Glasoe, G. N., and J. V. Lebacqz, Pulse Generators, McGraw-Hill Book Co., New York, 1948
31. Code X3
32. Reed, R. E., and A. C. Tunis, "Super-Power, Ultra-High-Frequency Amplifier Tube Developments," Presentation before AIEE General Winter Meeting, February 1961
33. Bennett, W. P., and H. F. Kazanowski, "One-Kilowatt Tetrode for UHF Transmitters," Proc. IRE, Vol. 41, p. 13, 1953
34. Salzberg, B., and A. V. Haeff, "Effects of Space Charge in the Grid Anode-Region of Vacuum Tubes," RCA Review, January 1938
35. Moreno, T., Microwave Transmission Design Data, Dover
36. Slater, J. C., Microwave Electronics, D. Van Nostrand Co.
37. Schelkunoff, S. A., Electromagnetic Waves, D. Van Nostrand Co.
38. Ramo, S., and J. R. Whinney, Fields and Waves in Modern Radios, Wiley, New York
39. Code X4
40. Weinberg, L., "Optimum Ladder Network," Jour. Franklin Institute, July 1957, August 1957
41. Mather, N. W., "An Analysis of Triple-Tuned Coupled Circuits," Proc. IRE, Vol. 38, No. 7, July 1956
42. Mather, N. W., and A. E. Harrison, "Graphical Analysis of Tuned Coupled Circuits," Proc. IRE, Vol. 37, No. 9, September 1949
43. Van Weel, A., "A Comparison of the Bandwidths of Resonant Transmission Lines and Lumped LC Circuits," Philips Research Report, Vol. 5, p. 241, August 1958
44. Terman, F. E., Radio Engineers Handbook, McGraw-Hill Book Co., New York, pp. 148-172
45. Reference Data for Radio Engineers, Federal Tele-

phone and Radio Corporation
46. Ginzton, E. L., Microwave Measurements, McGraw-Hill Book Co., New York, 1957
47. Benford, F., "The Blackbody," Parts I and II, General Electric Review, Vol. 46, Nos. 7 and 8, pp. 377-382, 433-440, July-August 1943
48. Temperature, Its Measurement and Control in Science and Industry, American Institute of Physics
49. Reed, R. E., "Final Technical Report Covering the Development and Refinement of the RCA Developmental Tube Type A2346," Vol. II, Contract AF30(602)-1396, RADC TPR-61-278, December 1961
50. Von Hippel, A. R., Dielectric Materials and Applications, Wiley, New York, 1954
51. Hansen, M., Constitution of Binary Alloys, Second Edition, McGraw-Hill Book Co., New York
52. Roark, R. J., Formulas for Stress and Strain, McGraw-Hill Book Co., New York, 1954
53. Timoshenko, S. P., Theory of Plates and Shells, McGraw-Hill Book Co., New York
54. Martin, I. E., J. A. Powell, and J. Zollman, "Ceramic, Sapphire, and Glass Seals for the Model C-Stellarator," Vacuum Technology Transactions Proceedings of the Sixth National Symposium, Pergammon Press
55. "Electrical Losses in Ceramic-to-Metal Seals," Univ. of Ill. Elec. Eng. Research Lab Report, February 1954
56. Carpenter, H., and J. M. Robertson, Metals, Vols. I and II, Oxford University Press, 1939
57. Metals Handbook, American Society for Metals, Cleveland
58. Palmer, F. R., and Luerssen, Tool Steel Simplified, Rev. Ed., The Carpenter Steel Co.
59. Seasholtz, A. P., Steel, Sept. 3-Oct. 1, 1945
60. Handbook of Chemistry and Physics, Chemical Rubber Publishing Co.
61. Woldman, N. E., Engineering Alloys, American Society for Metals
62. Hoyt, S. L., Metals and Alloys Data Book, Rein-Publi shing Corp., New York, 1943
63. Kohl, W. H., Materials and Techniques for Electron Tubes, Reinhold Publishing Corp.
64. International Critical Tables, Vol. II, p. 358 (properties of metals) McGraw-Hill Book Co., New York
65. International Critical Tables, Vol. V, p. 216, McGraw-Hill Book Co., New York
66. Forsythe, W. E., Smithsonian Physical Tables, Smithsonian Inst., Washington
67. Smithells, C. J., (ed.) Metals Reference Book, Vols. I and II, Interscience Publishers
68. International Critical Tables, Vol. VI, p. 109 (electrical conductivity), McGraw-Hill Book Co., New York
69. Code X5
70. International Critical Tables, Vol. II, p. 304 (insulation material), McGraw-Hill Book Co., New York
71. International Critical Tables, Vol. VI, p. 348 (magnetic properties), McGraw-Hill Book Co., New York
72. Code X6
73. Everhart, Lindlief, Kanagis, Weissly, and Siegel, Mechanical Properties of Metals and Alloys, U.S. Government Printing Office
74. Tucker, W.A., "Thermal Expansion Coefficients for 51 Metals," Product Engineering, March 1951
75. See Fig. 15, p. 55 of this book
76. Code X7
77. Cranberg, L., "The Initiation of Electrical Breakdown in Vacuum," Jour. Appl. Phys., Vol. 23, No. 6, p. 518, May 1952
78. International Critical Tables, Vol. V, pp. 238-247 (optical measures & temp.), McGraw-Hill Book Co., New York
79. Jones, H.A., and I. Langmuir, "The Characteristics of Tungsten Filaments as Functions of Temperature," General Electric Review, pp. 310-319, 354-361, 408-412, June-July-August 1927
80. Smithells, C. J., Tungsten, Third Edition, Chapman & Hall
81. Malter, L., and D. B. Langmuir, "Resistance Emissivities, and Melting Point of Tantalum," Phys. Rev., Vol. 55, p. 743, April 1939
82. Malter, L., and D. B. Langmuir, "The Rate of Evaporation of Tantalum," Phys. Rev., Vol. 55, p. 748, April 1939
83. Nickel and Nickel Alloys, The International Nickel Co.
84. Wise, E. M., and Schaefer, "The Properties of Pure Nickel," Metals and Alloys, Sept., Nov., Dec., 1942
85. OFHC Brand Copper, The America Metal Co., Ltd.
86. Wilkins, R. A., and E.S. Bunn, Copper and Copper Base Alloys, Mc-Graw-Hill Book Co., New York
87. Phosnic Bronze, A report published by the Chase Brass and Copper Co., Waterbury, Conn.
88. Platinum Metals and Their Alloys, The International Nickel Co.
89. "Engineering Properties of Inconel and Inconel X," Technical Bulletin T7, The International Nickel Co.
90. Heat Treatment and Physical Properties of Austenitic Chromium Nickel Stainless Steel, The International Nickel Co., 1947
91. Ruthberg, S., "Investigation of the Properties of Natural Mica for Use in Electron Tubes," U.S. National Bureau of Standards
92. Miller, G. L., Zirconium, Academic Press, New York
93. Symposium on Cleaning Electron Device Components and Materials, ASTM Special Technical Publication #246 presented in Philadelphia, 1958
94. Anderson, H. W., "Effect of Total Voltage on Breakdown in Vacuum," Electrical Engineering, Vol. 54, p. 1315, 1935, and Vol. 55, p. 830, 1936
95. Trump, J. G., and R. J. Van DeGroff, "The Insulation of High Voltage in Vacuum," Jour. Appl. Phys., Vol. 18, p. 329, 1947
96. RCA Electron Tube Division Radiation Safety Manual
97. Parker, W. N., and M. V. Hoover, "Gas Tubes Protect High Power Transmitter," Electronics, Vol. 29, p. 144, January 1956
98. Brubaker, W.M., "A Method of Greatly Enhancing the Pumping Action of a Penning Discharge," Proc. of theSixth Nat. Symposium, Vac. Tech. Transactions, Pergammon Press
99. Zaphiropoules, R., "Design Considerations for High-Speed Getter Ion Pumps," Proc. of the Sixth Nat. Symposium, Vac. Tech. Transactions, Pergammon Press

Gas Tube Design

H. H. Wittenberg

Lancaster

The subject of gas tubes is an old one. Early electron tubes were gassy due to inadequate pumping techniques. Certain early amplifier tubes, on the other hand, were gas-filled in an attempt to use the gain resulting from avalanche ionization; early cathode-ray tubes also were gas-filled to assist focusing. The gas tube field enjoyed a heyday in the late 1920's when the phanotron (low-pressure gas diode), tungar (high-pressure gas diode), thyratron, and ignitron appeared in quick succession. The gas-tube field was quiescent during the thirties while vacuum tube development surged ahead. During World War II the transmit-receive (TR) box and the hydrogen thyratron made radar practical and revived interest in the gas-tube field. In the early 1950's, the gas-tube field received another "shot in the arm" with the developments of the Plasmatron and Tacitron by our Princeton Laboratories. At this writing the fusion process for releasing nuclear energy again focuses attention on gas tubes.

The theory upon which the foregoing devices are based exhibits a similar chronology. Townsend originated the early theory in 1915. Langmuir, Hull, and Slepian developed, in about 1929, the basic theory which is currently found in the text books. Because much of the phenomena are not yet understood, contemporary researchers, including S. C. Brown, S. T. Martin, L. Malter, and E. O. Johnson, using modern techniques, are now developing new and more comprehensive theory.

The gas tube is often thought of as being limited to the audio range in its frequency of operation; on the contrary, some gas tube phenomena occur at megacycle rates. TR boxes involve the radar frequencies, measurement techniques utilize ultrahigh frequency (UHF), and the yet unharnessed oscillations inherent in a gas discharge extend into the VHF spectrum.

The complexity of the gas tube may be ascribed to the fact that three types of particles are involved: molecules, ions, and electrons. Further complications arise from the varieties and combinations of gases which may make up the molecules and ions.

Although the gas-tube designer is spared the necessity of mastering certain vacuum-tube operational characteristics, he must be familiar with many techniques and characteristics of the vacuum-tube field such as pumps and exhaust, gettering, heaters, emission and coatings, anode materials, heat radiation and conduction, grid emission, and amplification factor. In addition, he must master the techniques of gas filling and understand the complex phenomena resulting therefrom. The scope of the material in this article is restricted to phenomena unique to gas tubes.

PROCESSES PRIOR TO CONDUCTION

Collision

Mean-Free-Path. Gas molecules may be ionized by electron impact. This process is of major importance since, when coupled with secondary emission of electrons from the cathode, it makes possible a self-sustaining gas discharge. An electron injected into a gas will collide with the gas molecules. If the molecules were spaced in an orderly array, such that along a given straight line the spaces between molecules were equal and the electron continued along this same straight line between collisions, then the free path between collisions would be a constant length. But the molecules are in motion and the electron direction is random and subject to change after every collision. Consequently, the free path of a given electron is continually changing.

The mean of this variable-length path can be found in the following manner (see Fig. 1). In traveling a free path through the gas, the electron may be imagined to cut a cylindrical path of cross-sectional area πr^2, where r is the molecular radius. If the center of the molecule lies within this path, it will be hit when the electron has traveled a length λ, called the mean free path.

When the unit volume of the path of the electron is equated to the unit volume per molecule:

$$\lambda \pi r^2 = 1/N \qquad (1)$$

where N is the number of molecules per unit volume.

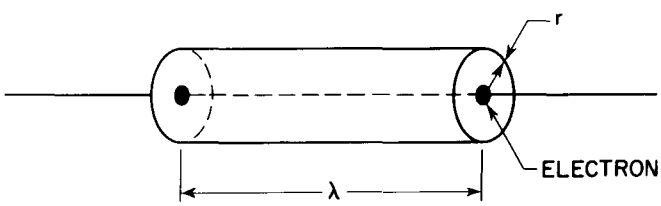


Figure 1. Mean Free Path of Electrons

The molecular mean free path, derived in a similar manner, is found to be 1/4 of the electron mean free path.[1] Because molecular speed is much lower than electron speed, the second or target molecule will have moved during the free flight of the first molecule. A more rigorous calculation taking this thermal motion into account gives the molecular free path,

$$\lambda_m = 1/4\sqrt{2}\ \lambda e \qquad (1a)$$

where λe is the electron mean free path.

The molecule collides more frequently per given distance of travel.

By the gas laws:

$$N = PV/kT \qquad (2)$$

where P = pressure (dynes/cm^2)
V = volume (cm^3)
T = temperature, degrees Kelvin
k = Boltzmann's Constant (1.38×10^{-16} erg/deg)

Hence, the mean free path is proportional to the absolute temperature of the molecules and inversely proportional to the pressure of the molecules. Table I gives the mean free path of the gases used in gas tubes at a typical dosing pressure.[2]

TABLE I. Mean Free Path

Gas	Mean Free Path in millimeters*	
	For Molecule	For Electron In Gas
H_2	0.90	5.1
He	1.41	8.0
Ne	1.00	5.66
A	0.51	2.88
Kr	0.39	2.2
Xe	0.26	1.47
Hg	0.35 (100 μ, 82 C)	1.98
	3.5 (10 μ, 47 C)	19.8
	35.0 (1 μ, 18 C)	198.0

*Note: At 0 C and 100 microns (μ) pressure except as indicated

Free Paths vs. Number of Particles.[3] Because the motion of gas molecules is random, the free paths are randomly distributed with regard to distance. Some paths are long while others are short. Since λ is the mean free path of the electron, $1/\lambda$ is the mean collision frequency (collisions per centimeter) per electron. The number of electrons dn colliding while traversing a distance dx is:

$$dn = -n \frac{1}{\lambda} dx \qquad (3)$$

where n = number of electrons traveling distance x without collision

The negative sign signifies the decrease in n as x is increased.

Integrating Eq. (3) and substituting boundary conditions gives:

$$n = n_o e^{-x/\lambda} \qquad (4)$$

where n_o = total number of electrons

This is known as the survival function and describes the number of electrons surviving in their original path without collision after traveling distance x. It is more valuable to know the number of electrons having free paths in the range of x to (x + dx). Differentiating Eq. (4) with respect to x gives:

$$\frac{dn}{dx} = \frac{n_o}{\lambda} e^{-x/\lambda} \qquad (5)$$

To use this relationship, the differentials are replaced by increments and regrouped to give

$$\frac{\Delta n}{n_o} = \frac{x}{\lambda} e^{-x/\lambda} \qquad (6)$$

Eq. (6) is plotted in Fig. 2. Note that in this case the distribution function of Eq. (6) is the same as the survival function of Eq. (4).

To better appreciate these relationships, the distribution function of men's heights may be plotted, as in Fig. 3, corresponding to the function of Eq. (4), but having a different shape. If knowledge of the number of men

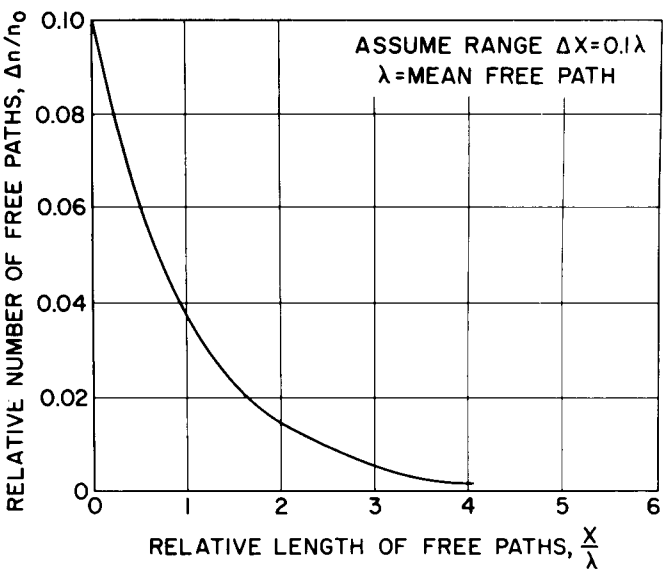


Figure 2. Distribution of Free Paths

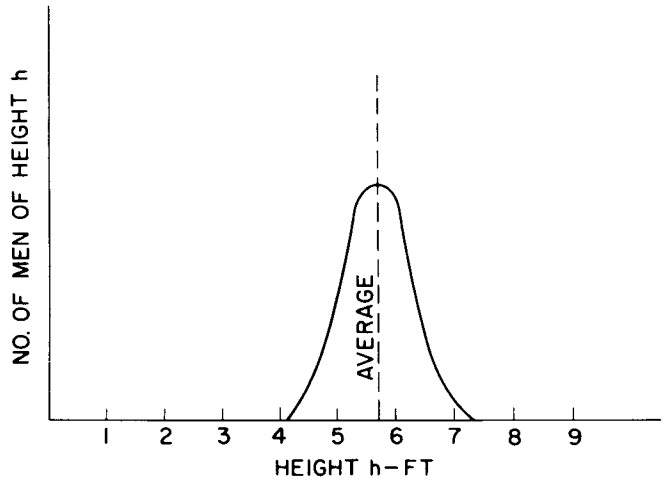


Figure 3. Distribution of Men's Heights

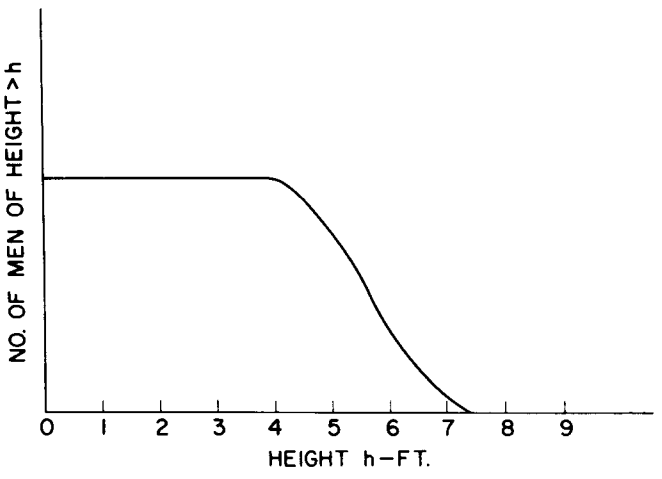


Figure 4. Integrated Distribution of Men's Heights

with heights greater than a given value h is desired, Fig. 3 may be integrated to get Fig. 4. Fig. 4 is the survival curve corresponding to the function of Eq. (4).

Examination of Fig. 2 shows that 10 per cent of the free paths are in the range of 0 to 0.1 of the mean free path λ. Very few electrons have paths greater than λ; only 3.7 per cent have free paths in the range centering at λ.

Free Paths vs. Voltage. If the electrons and molecules are considered as solid spheres, it would appear that the mean free path were independent of speed of the moving particles. However, if the molecule is pictured as a planetary system of electrons, it is easy to imagine that a fast moving electron may miss the planetary electrons and so avoid collision. Experimental evidence shows that the mean free path is, indeed, a function of speed S. It is customary to express this function[4] in terms of the mean collisions per unit distance P_c and voltage V (see Fig. 5). The reciprocal of P_c is the mean free path. This quantity P_c has been miscalled probability of collision, but is not a probability since it has dimension 1/centimeter. From Eq. (1) it can be seen that $1/\lambda$ is proportional to the cross section and hence is sometimes expressed as collision cross section.

Voltage is related to speed by the formula

$$S = \left(\frac{2qV}{m}\right)^{1/2} \text{ meters/sec} \qquad (7)$$

where q = charge of particle in coulombs
m = mass of particle in kilograms

The arrows in Fig. 5 show the kinetic-theory values which were calculated using only the speed of thermal motions (equivalent to fractions of a volt).

Ionization

Voltage Dependence. Having collided with an atom, the colliding electron may affect the planetary electrons

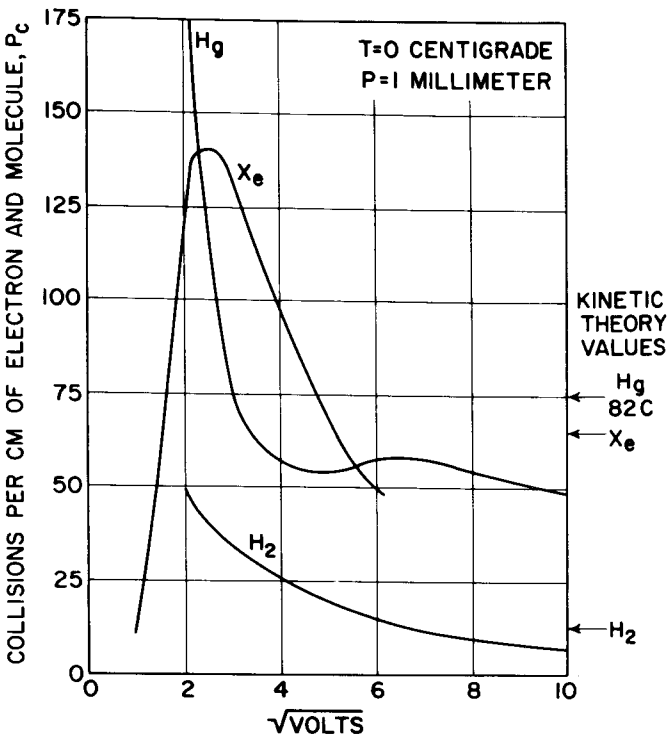


Figure 5. Probability of Collision

in the following ways:

(a) A planetary electron may be excited, remain so for about 10^{-9} second, and then radiate the energy received in the collision.

(b) A planetary electron may be excited, remain so for 10^{-3} second (metastable), and suffer a second collision from which it may be removed from the atom (cumulative ionization).

(c) A planetary electron may be removed directly (ionization).

(d) The electron may simply bounce off like a pea from a cannon ball.

Table II gives critical voltages at which effects described in (a) through (c) begin to occur.

Probability of Ionization. Like the probability of collision, the probability of ionization actually is not a probability but a coefficient. It is expressed as ions per centimeter per colliding particle. (Townsend labeled this coefficient α.) The ionizing coefficient is shown in Fig. 6 as a function of voltage.[5] Above the threshold values (shown in Table II), the probability of ionization increases to a peak and then falls. It will be noted that the ionization coefficient is maximum at about five times the threshold ionization voltage.

BREAKDOWN

Secondary Emission at Cathode

Assume that two parallel plane electrodes, immersed in a gas, have a voltage applied to them. Stray radiation striking the negative electrode will release a small number of electrons. The electrons move toward the positive electrode and further ionize in a manner just described. Likewise, the ions move toward the negative electrode colliding frequently with molecules, but few ions are produced in such collisions.[6] However, upon hitting the negative electrode the ions are effective in releasing electrons by the process of secondary emission (Townsend labeled this coefficient of this process γ.). Fig. 7 is a graph[7] of the secondary emission vs. the electric field (in volts per centimeter) for hydrogen

TABLE II. Critical Voltages of the Gases[17,18]

Gas	Weight	Vapor Pressure*	First Ionization Voltage Volts	Second Ionization Voltage Volts	Excitation Voltage Volts	Metastable Voltage Volts
Cs	55	5×10^{-4}	3.9	23.4	1.4	
Rb	37	1×10^{-4}	4.2	27.4	1.5	
K	19	2×10^{-5}	4.3	31.7	1.6	
Na	11	1×10^{-7}	5.1	47.1	2.1	
Li	3	10^{-8}	5.4	75.3	1.8	
Ca	20	10^{-8}	6.1	11.8	2.5	1.9
Mg	12	10^{-8}	7.6	15.0	4.3	2.7
Hg	80	3×10^{-1}	10.4	18.1	4.9	4.7, 5.4
Xe	54		12.1	21.1		8.3, 8.9
H_20	18		13.2		17.6	
H	1		13.4		10.1	
Kr	36		13.9	26.4		9.9, 10.5
O_2	32		14.1		6.1	
CO	28		14.2		6.0	3.5
CO_2	44		14.3		10.0	
H_2	2		15.3		11.1	
A	18		15.7	27.8		11.5, 11.7
N_2	28		16.7		8.5	
Ne	10		21.5	40.9		16.6, 16.8
He	2		24.5	54.1		19.7, 20.5

*Note: Vapor pressure in mm Hg at 100 C

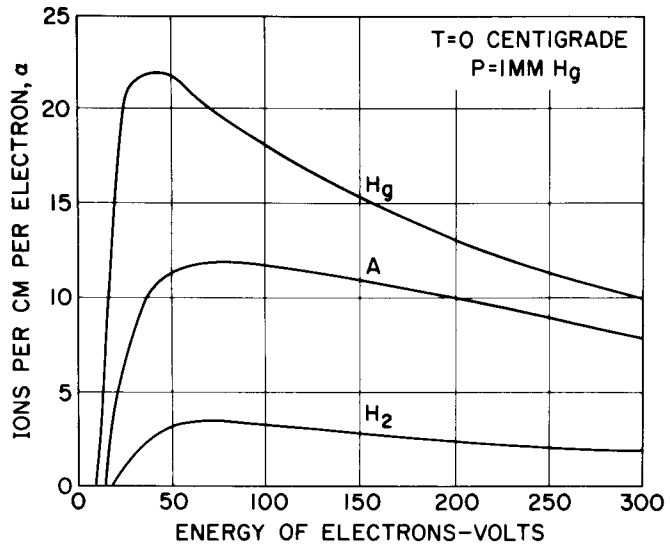


Figure 6. Probability of Ionization

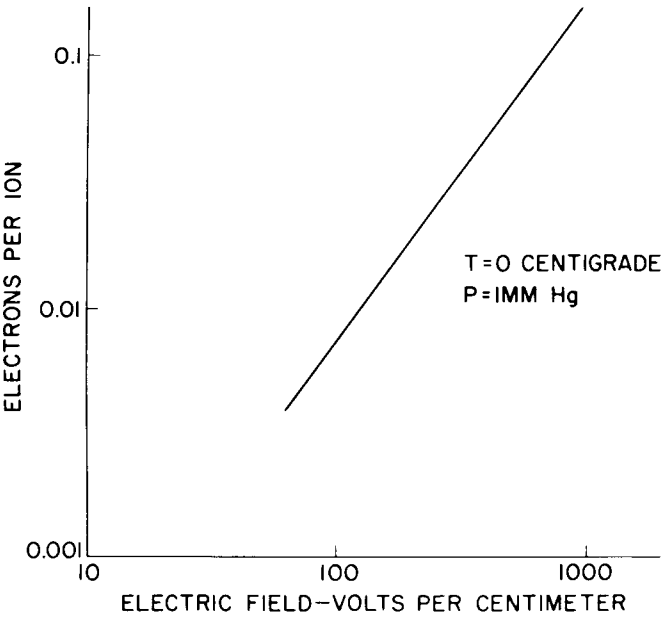


Figure 7. Secondary Ion Emission

with zinc electrodes. If the two processes of ion production by electron collision with a molecule and electron production by ion impact at the cathode are combined, the elements of a feedback system are perceived to be present. If the gain of the system exceeds one, the result is breakdown of the gas. Limitations in the form of external resistance and ion losses with the gas-filled envelope prevent unlimited current build-up and stabilize the gas discharge in various modes to be described later. See Appendix A for a mathematical treatment of current build-up.

Paschen's Law

The voltage at which the ion-electron generation system feedback becomes unstable is called the breakdown voltage. It is common experience that at atmospheric pressure the breakdown voltage decreases as the distance between electrodes is decreased. It is not common knowledge that at low gas pressures the distance can be decreased to a point where the breakdown voltage reaches a minimum and thereafter increases as the distance is further decreased. The complete relationship between breakdown voltage, distance, and pressure was first discovered by F. Paschen in 1889 and is known as Paschen's Law.

Paschen's Law may be deduced in a nonmathematical fashion. Referring to Fig. 8, the line labeled "Collisions per Electron in Traversing X_a" rises with increasing pressure because the gas density increases with pressure and consequently there are more molecules inserted in the path of the traveling electron. The line labeled "Energy per Collision" falls as pressure increases, because (assuming that the electron loses all of its energy at each collision) the electron has a shorter distance between collisions to pick up energy from the electric field. As was shown previously, because a certain energy must be acquired to produce ionization on impact, the total number of ions produced must be a product of these two values. This product is shown as the curve "Ions Produced". It has a maximum near the crossover of the two lines. The more ions that are produced, the less voltage is necessary to cause breakdown. Hence, the breakdown voltage curve has a minimum. A similar reasoning process can be carried through for distance as a variable, with pressure constant. In the complete Paschen's Law, breakdown voltage is a function of the product of distance and pressure (pd). A plot of Paschen's Law exhibits the same shape as the curve in Fig. 8. Plots of Paschen's Law for the various gases, have different minimums as Fig. 9 shows, but all exhibit the same characteristic shape. A mathematical derivation of Paschen's Law is given in Appendix B.

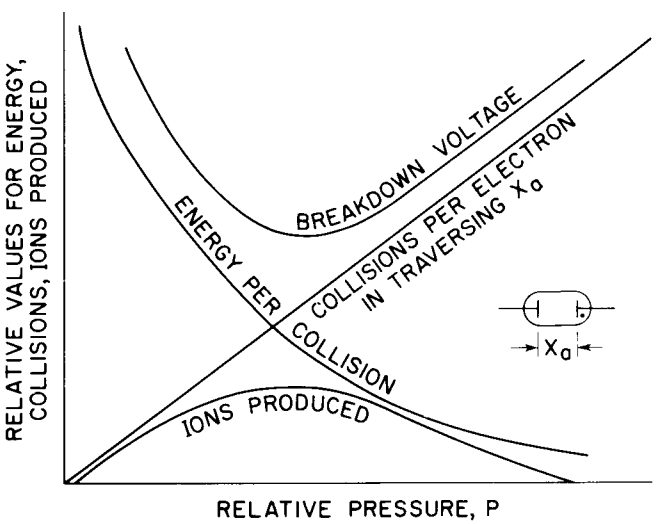


Figure 8. Non-mathematical Derivation of Paschen's Law

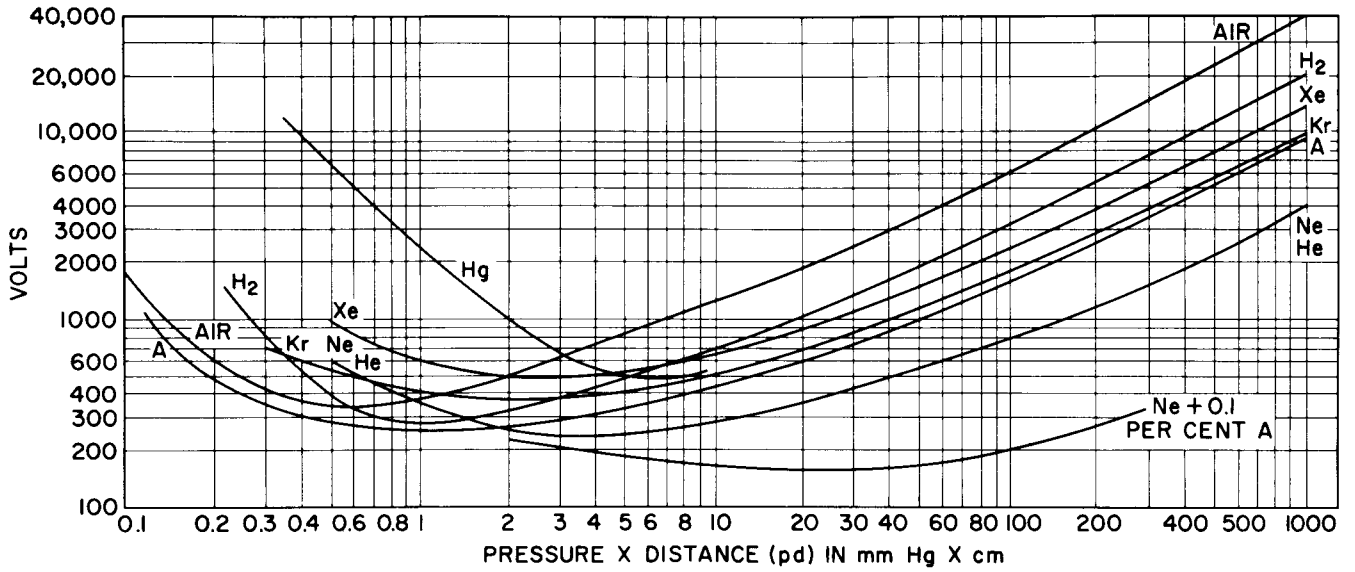


Figure 9. Paschen's Law for Various Gases

PROCESSES DURING CONDUCTION

Voltage vs. Current

Fig. 10 shows the voltage-vs.-current relationship for a gas discharge. As the voltage is increased from zero, a small current called the Townsend current flows. Ionization gauges operate in this region and the gas test for vacuum tubes is performed here. With irradiation of the cathode, the curve shifts to higher currents as shown by the dotted curve. Gas phototubes are operated in this region and realize the benefits of gas amplification. Geiger-Mueller (GM) tubes also operate in this region and indicate the presence of cathode irradiation by current increases. At further increases in supply voltage, a plateau is encountered; such a flat characteristic permits the use of a gas-filled tube as a corona-type gas voltage regulator. Although in Fig. 10 the plateau is at about 400 volts, different gases and different electrode materials can shift this region upward to several kilovolts.

On the plateau just described breakdown occurs which, without the limitation of a series resistance, will result in an uncontrolled increase of current. Large series resistance and a large supply voltage permit the negative resistance portion to be traversed. Farther along the solid line is the glow-discharge region. In this region the glow-discharge voltage regulators (VR tubes) and the transmit-receive switches (TR tubes) are operated. Cathodes of such tubes are relatively cold and emission occurs by virtue of ion bombardment.

Further increases of current lead to the sudden drop of terminal voltage to about 10 volts. A very low voltage and an incandescent cathode are characteristic features of the arc discharge. If the cathode is externally heated, the glow discharge mode is bypassed and the low voltage region is entered immediately. It is in this region that the thyratrons and hot-cathode gas rectifiers operate.

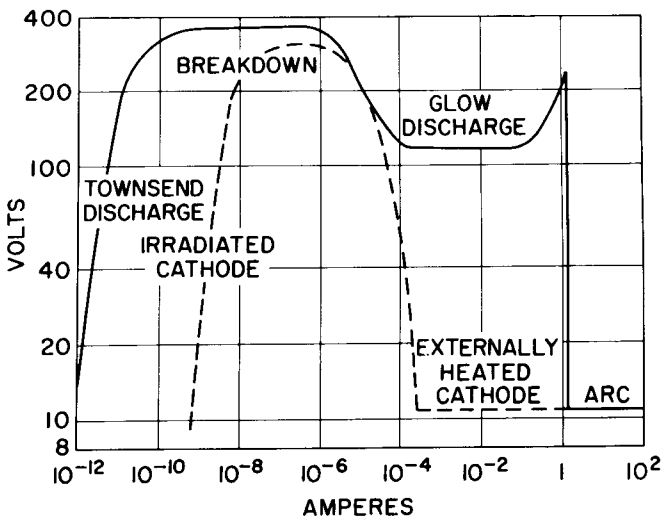


Figure 10. Voltage vs. Current

Voltage vs. Distance

In a vacuum diode having parallel plane electrodes, the voltage[8] is proportional to the 4/3 power of the distance between cathode and anode; the slight deviation from linearity is a result of the negative electron space charge. In a gas diode having parallel plane electrodes, the voltage distribution with distance may take several modes depending upon (1) whether the negative-electron or the positive-ion space charge predominates and (2) upon boundary conditions. Fig. 11 shows three of the voltage distributions[9] which have been identified. The Langmuir mode is characterized by a sharp rise within about a millimeter of the cathode and a practically constant voltage from there to the anode. Visually, the region next to the cathode (identified as the "sheath") is dark, indicating the absence of excited atoms and ions. The remainder of the volume (identified as the

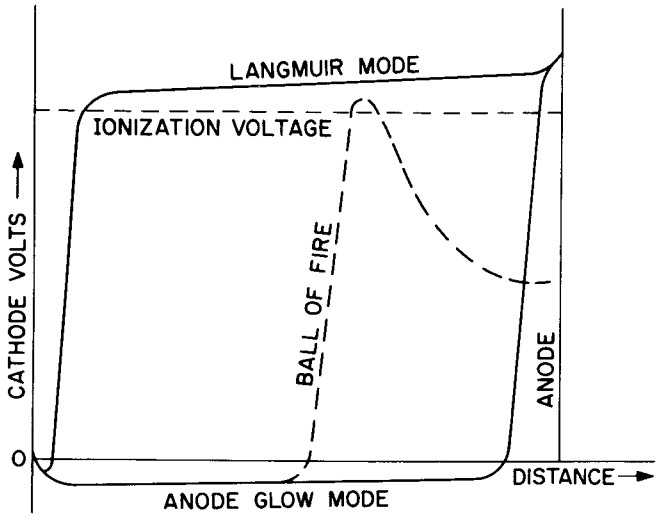


Figure 11. Voltage vs. Distance

"plasma") glows indicating the presence of excited atoms and ions. The sheath draws electrons from the cathode and accelerates them to sufficient velocity for ionization. The existence of this sheath makes possible the design of gas-tube cathodes with several times the thermal efficiency of vacuum-tube cathodes.

The Langmuir mode is most common in typical gas tubes. When the anode current in a typical gas tube is reduced to a small percentage of the available emission (as determined in a vacuum) the anode-glow mode[10] may be observed. As Fig. 11 shows, the sheath occurs close to the anode. The electrons enter the field-free region from the cathode as a result of initial velocities. The anode-glow mode may be enhanced by increasing pressure above 300 microns and by minimizing ion losses. The inverted structure in which the cathode surrounds the anode is capable of operating in the anode-glow mode to higher anode currents than conventional structures. The Tacitron, designed to operate in this mode, exhibits advantages of low noise and current cutoff ability.

At current levels intermediate between the Langmuir mode and the anode-glow mode, the ball-of-fire mode is encountered. As may be seen in Fig. 11, the voltage rises to a maximum as the point of observation progresses across the tube. With such a distribution the maximum voltage may be sufficient to ionize whereas the over-all voltage across the tube may be appreciably less.

Sheath

A sheath in a gas tube is defined as a region exhibiting appreciable electric field. Its thickness is small compared to the mean free path so that electrons and ions cross the sheath without collision. In the sheath just described, electrons emitted from the cathode enter the sheath on one side, whereas ions emitted from the plasma enter the sheath on the other side. Such a sheath is actually called a double sheath. The single sheath, which involves charges of only one kind, is the simplest mathematically and so will be considered first. The mathematical solution to the single sheath case is identical to the solution of the space-charge-limited vacuum-diode case, known as Child's Law.[11]

$$i = 5.23\ 10^{-8}\ M^{-1/2}\ d^{-2}\ V^{3/2} \qquad (9)$$

where i = current in amperes per square centimeter
M = atomic weight of ion
d = thickness of sheath in centimeters
V = voltage across sheath in volts

This formula holds for the case of an electrode immersed in a plasma. Application of a voltage to such an electrode causes the sheath to be formed. The plasma serves as the emitter of positive ions which move across the sheath to the electrode. The density of the current to the electrode is identical to the density of the current across a surface in the plasma (assuming the electrode to be large with respect to the sheath thickness). It will be seen later that the plasma current density is fixed by the conditions of the plasma and, to also satisfy Child's Law, the thickness of the sheath adjusts itself. A double sheath exists when electrons enter the source from one side and ions enter the sheath from a source on the other side. Consider the sheath to occupy the entire space between cathode and anode, which means that electrons cross the sheath and collide with gas molecules directly in front of the anode. A mathematical solution to this double sheath problem[12] shows that the current in the external circuit is 86 per cent greater for the gas-filled device than for the vacuum device. This increased current consists of 98 per cent electrons and 2 per cent ions. The ions contribute little because they move slowly, but they reduce the negative space charge permitting the larger electron flow.

If the circuit voltage is increased, the current increases and the sheath thickness decreases to less than the anode-to-cathode distance. A plasma develops in the vacated space between the sheath and the anode. Because the plasma is a region of high conductivity, this case may be described as one in which a virtual anode exists at the sheath boundary. In comparison, the current flow for the vacuum diode would increase with a decrease of anode-to-cathode distance. Mathematical considerations show that the current for the gas-filled device is always 86 per cent greater than that for the vacuum diode if the anode in the vacuum diode is placed at the same distance from the cathode as the virtual anode is in the gas diode. In actual tubes, this procedure is impractical because the sheath generally reduces across to a millimeter or less.

Plasma

Description. The plasma may be defined as a region of very small electric field. The electron and positive ion densities are approximately equal. For example, in mercury at a pressure of 1 micron, a discharge of 170 milliamperes per square centimeter has 10^{11} ions per cubic centimeter, 10^{11} electrons per cubic centimeter, and 4×10^{13} atoms per cubic centimeter. Note that only one out of 4000 atoms is ionized. The atoms move about as a result of their thermal energy which is relatively low[13] (0.04 volts at 25 C).

Recombination of an electron and an ion in the plasma is rare because the electron is unable to dispose of its kinetic energy in the collision with the heavier ion. See Appendix C for the mathematics of this process.

Ions and electrons can combine easily at the walls of a discharge tube. When a particle of either kind strikes a wall, it gives up its kinetic energy as heat and adheres by electrostatic attraction until a particle of the opposite sign arrives to form a neutral atom.

Velocity Distribution. To analyze the behavior of the plasma it is necessary to determine the velocity distribution of the particles therein. In the literature[14] may be found the theory which culminates in the following function:

$$f(V_x^2) = A_1 e^{-A_2 V_x^2} \tag{10}$$

where V_x = component of velocity in x-direction
$f(V_x^2)$ = fraction of particles having speeds in range dV_x divided by dV_x

$$A_1 = \left(\frac{m}{2\pi KT}\right)^{1/2}$$

$$A_2 = \frac{m}{2KT}$$

m = mass of particle
K = 1.38×10^{-23} joule per degree
T = temperature in degrees Kelvin

This function is plotted in Fig. 12 for hydrogen. The curves show that the most probable velocity component in the x-direction is zero, but the resultant velocity is not zero because the velocities V_y and V_z in the other directions probably are not zero. Note that the distribution is a function of V_x^2 and is, therefore, the same in both the positive and negative directions. As the temperature is increased, the number of particles having higher velocities increases, but the area under the curve (the total number of particles in the given volume) remains constant.

The temperature of the ions is about the same as that of the molecules of the gas (a few hundred degrees Kelvin); both types of particles collide among themselves and with each other, and thereby interchange energy. The electrons are found to have a much higher temperature than the ions or molecules (in the tens of thousands of degrees Kelvin).

Particles Crossing Plane.[15] To determine the number of particles crossing a sheath, it is necessary to know how many particles cross a unit area (ABCD of Fig. 13) in one direction. All of the particles in the box of height V_x and base ABCD, having velocities in the range V_x to $(V_x + dV_x)$, will cross area ABCD in one second. Of course, some of the particles starting at the far end of the box will leave the box because of y and z components of velocity, but they will be replaced by an equal number of outside particles. The total number of particles in the box is given by

$$n_v = N \cdot V_x \cdot 1 \tag{11}$$

where N = density of particles

Utilizing the velocity distribution function of Eq. (10) to obtain the number of particles in the box having velocities between V_x and $(V_x + dV_x)$ gives

$$dn_v = NV_x A_1 e^{-A_2 V_x^2} \tag{12}$$

where the constants A_1 and A_2 are functions of the temperature and of the gas only

To find the total number of particles crossing the unit area per second, it is necessary to integrate over all velocities (0 to ∞) giving

$$n_{0,\infty} = NA_1 \int_0^\infty V_x e^{-A_2 V_x^2} dV_x \tag{13}$$

or

$$n_{0,\infty} = \frac{NA_1}{2A_2} = \left(\frac{KT}{2m\pi}\right)^{1/2} \tag{14}$$

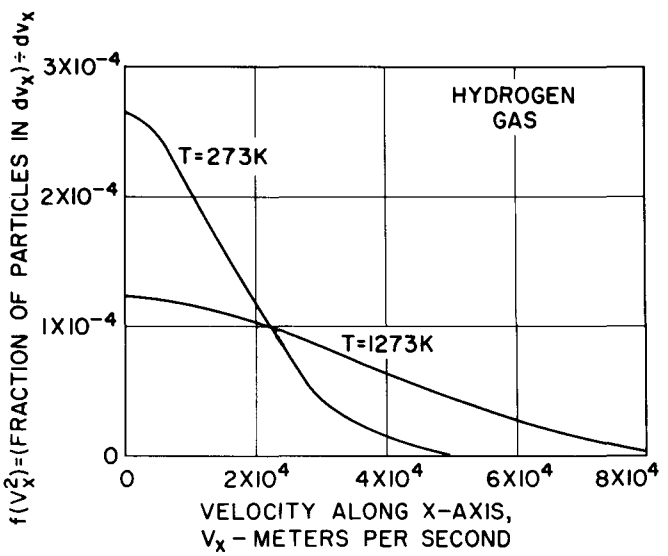


Figure 12. Velocity Distribution (Hydrogen gas)

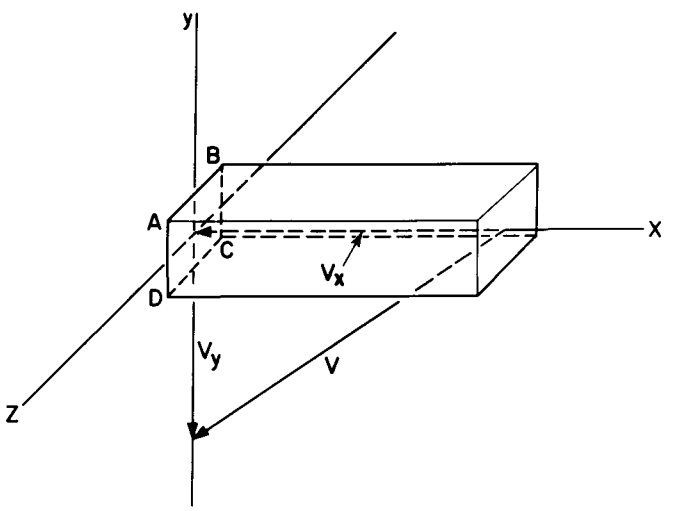


Figure 13. Particles Striking Wall

Multiplying the number of particles crossing the unit area per second by the charge per particle gives the random plasma current density J_r. Useful in the analysis of sheaths is the expression for the number of charged particles crossing the unit area against a retarding voltage. The particle must have a kinetic energy $(1/2mV_x^2)$ equal to or greater than the potential energy qE (where q is the charge and E the voltage):

$$1/2mV_x^2 = qE \tag{15}$$

Substituting V_x for the lower limit instead of zero and integrating gives

$$n_{V,\infty} = n_{0,\infty}\, e^{\frac{-mV_x^2}{2KT}} \tag{16}$$

$$= n_{0,\infty}\, e^{\frac{-qE}{KT}} \tag{17}$$

Probes. The insertion of a third electrode in a gas discharge between cathode and anode sets the scene for a situation known as "probe behavior." If the voltage of the third electrode or probe is varied with respect to the cathode, the current will vary in the manner shown in Fig. 14. When the voltage of the probe is more than a few volts below that of the cathode (line A-B), the current is small and at saturation level; the direction of the current and the saturation shows that the probe is collecting positive ions only. Similarly, along line C-D, the probe is collecting only electrons. These saturation effects are accompanied by sheaths which surround the probe and expand (see Eq. (9) and discussion on sheaths) as the voltage between the probe and the plasma increase.

Along line B-C, the probe is collecting both positive ions and electrons. When the probe current is zero, the numbers of ions and electrons collected are equal. The corresponding voltage F is not the space voltage, but is some negative value since the faster moving electrons must be repelled to equalize the numbers of both particles collected.

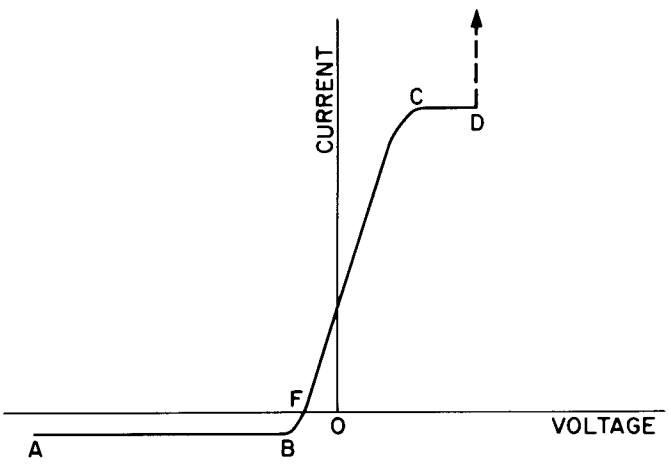


Figure 14. Probe Characteristic

As long as the probe voltage is less than the space voltage, the positive-ion current remains constant. Starting at B and making the voltage more positive, the electron current I_e increases according to the following modification of Eq. (17)

$$I_e = n_{0,\infty}\, Sq\, e^{\frac{-qE}{KT_e}} \tag{18}$$

where S = area of probe
q = charge
T_e = electron temperature
E = voltage difference between probe and space

Taking the logarithms of Eq. (18)

$$\ln I_e = \ln A_3 - \frac{q}{KT_e} E \tag{19}$$

When the experimental data are substituted in Eq. (19) and plotted, the straight-line portion gives the electron temperature T_e. The voltage at which the plot departs from a straight line is considered to be space voltage. At a more positive voltage, the positive-ion space charge disappears and a negative electron space charge appears. With T_e determined and assuming saturated electron current, Eq. (14) can be used to obtain the electron density N_e. The positive ion density is $N_p = N_e$.

Although a value for ion temperature can be determined by applying Eq. (14) to ions, the value thus determined is higher than the gas temperature, which is considered to be the correct value for ion temperature.

RECTIFIERS

Anode Efficiency

The efficiency of the hot-cathode gas rectifier accounts for its nearly exclusive use in the application range of 1000 to 20,000 peak inverse volts. For average currents from 0.1 ampere to 10 amperes, the gas tube displays a low tube drop (about 10 volts), whereas the drop across an equivalent vacuum tube is high (about 1000 volts). The resultant circuit efficiency can be as high as 96 per cent for gas tubes compared to 60 per cent or less for vacuum types. Average anode currents for gas rectifier tubes range from 0.1 to 40 amperes for the hot-cathode type, and from 5 to 400 amperes for the pool-cathode type. (The designation "hot" refers to a temperature at which radiation from the cathode is visible; such a temperature normally is attained by the application of external heat. The voltage-current operating region of the externally heated cathode type is shown in Fig. 10. A hot cathode may be heated internally by ions. However, the number of types using this principle (excluding fluorescent lamps) is negligible. A gas discharge tube in which the cathode is a pool of a liquid metal (usually mercury) is classified as a cold-cathode type of special form.) A further advantage of the gas rectifier is the superior regulation of output voltage obtainable from power supplies employing gas rectifiers. This good regulation results from the constancy of tube drop from no-load to full-load current.

The design of tubes using pool cathodes will not be considered since such types are not manufactured by RCA. However, the RCA product line does include purchased mercury-pool types.

The hot cathode of a gas tube may be either directly or indirectly heated; both types are employed in tubes presently manufactured. The directly heated cathodes have the advantages of fast warm-up time, simplicity, and low cost. The indirectly heated cathodes have the advantages of high thermal efficiency and freedom from phase effects and hum effects.

Cathode Efficiency

The hot cathode of a gas tube offers a much higher thermal efficiency than its vacuum counterpart because in a gas tube the anode current may be drawn from crevices and holes in the cathode surface. This phenomenon occurs because the plasma extends to within about a millimeter of the cathode surface. Thus, the emitting surface can be convoluted so that thermal radiation is conserved but electrons may emerge via the conductive plasma. Experience has shown that 1100 K is the optimum temperature for the operation of the oxide-coated cathode in a gas tube. To maintain a planar cathode at 1100 K with an emissivity of 0.25 (a value taken for barium oxide) requires 2.1 watts per square centimeter, assuming no conductivity losses. Conductivity losses and poor thermal coupling of heater and cathode may run the required power to as high as 7 watts/cm^2 for a cylindrical outside-coated cathode. Convoluting the cathode surface, reducing losses, heat shielding, and good thermal coupling to the heater may reduce the power required for 1100 K operation to as low as 1 watt/cm^2. The indirectly heated cathode of type 3C45 (6130) (see Fig. 15) has a good thermal efficiency (3.6 watts/cm^2) due to the coated insert. The indirectly heated cathode of the type 5C22 (see Fig. 16) has excellent thermal efficiency (1.56 watts/cm^2) due to the fins and multiple heat shielding. Although the directly heated cathode has perfect thermal coupling between heater and cathode, other factors may prevent the attainment of high thermal efficiency. The directly heated cathode of type 5557 (see Fig. 17) (3.45 watts/cm^2) does not realize optimum efficiency because most of its area can radiate to external cooler surfaces (no convolutions). The directly heated cathode of type 866A (see Fig. 18) is an improvement due to the edgewise winding of the ribbon (2.55 watts/cm^2). The directly heated cathode of type 5855 (Fig. 19) (3.45 watts/cm^2) has a high thermal efficiency due to the inside coating and heat shielding.

Cathode Design

Average Current. When a gas rectifier or thyratron is to be designed, the service to which it will be applied determines the average and peak cathode currents. The cathode area is a function of these currents, the type of emitting surface, and the expected life. Most gas tubes use the conventional oxide emitting surface and 2000 hours may be considered a medium life expectancy. Experience has shown that, for gas tubes using the oxide emitting surface, the average cathode-current density should not exceed 0.200 ampere per square centimeter of cathode area for a life expectancy of 2000 hours; a longer life expectancy requires a lower current density.

Peak Current. Consideration of peak current may require modification of the cathode area as determined by the above figure. Peak current, as used here, means the current having a duration as represented by half a sine pulse from a 60-cycle-per-second source and hav-

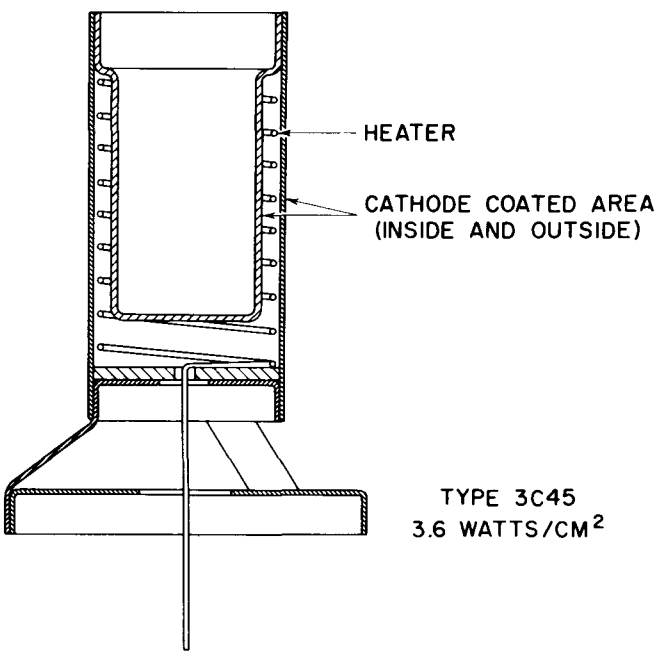


Figure 15. Indirectly Heated Cathode Structure of Type 3C45

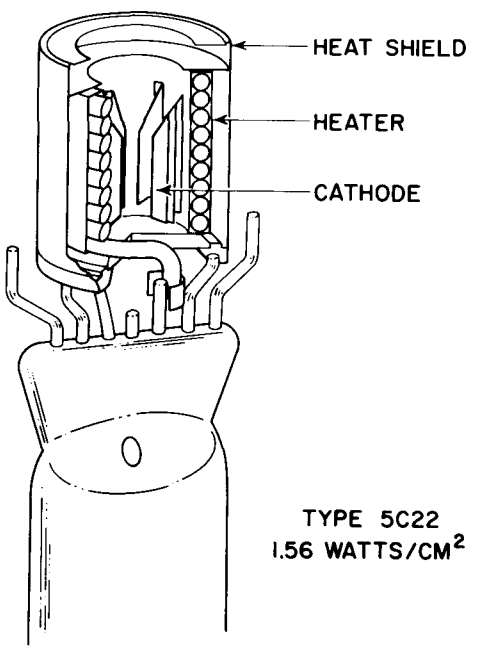


Figure 16. Indirectly Heated Cathode Structure of Type 5C22

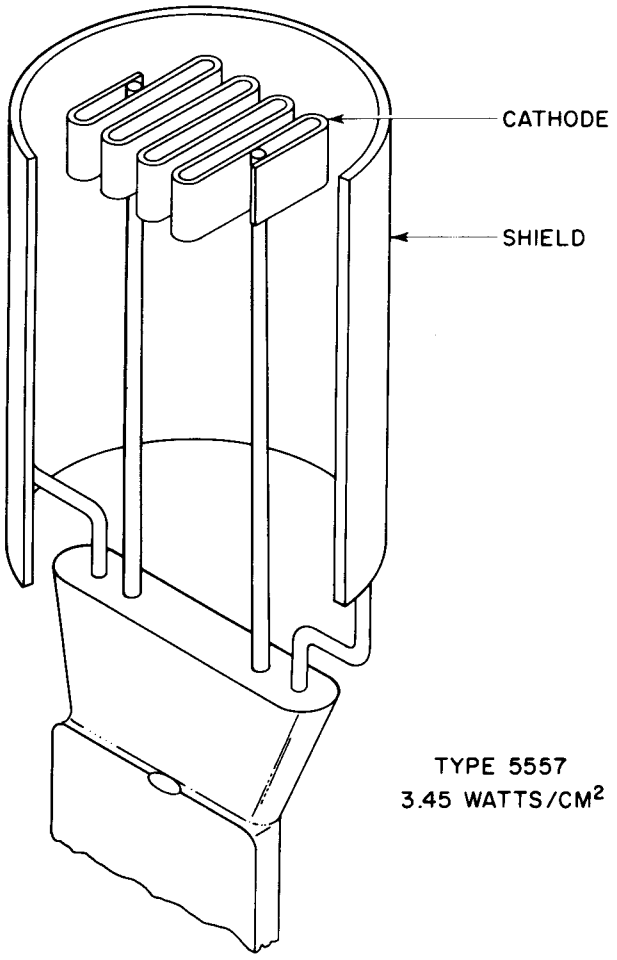


Figure 17. Directly Heated Cathode Structure of Type 5557

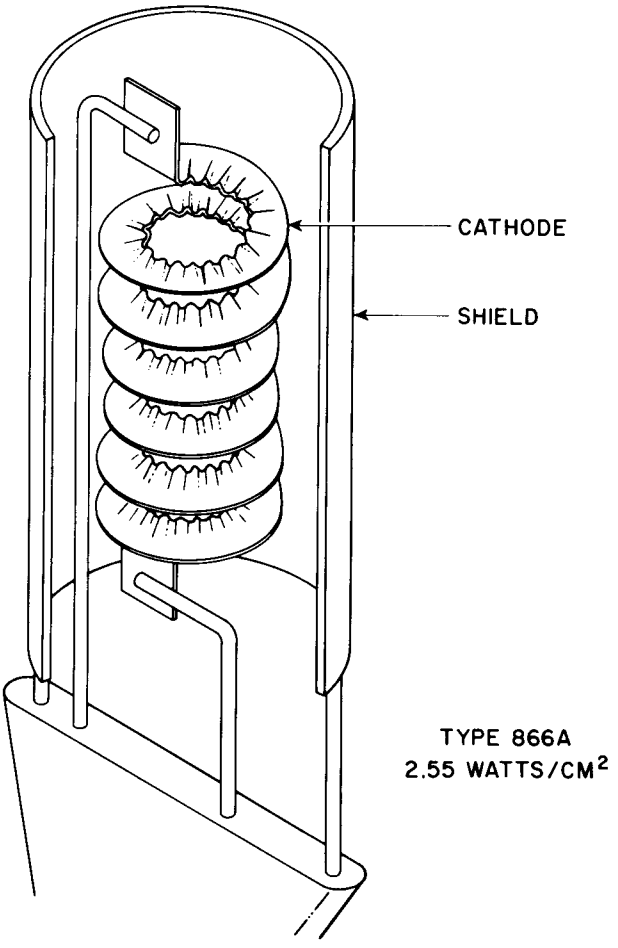


Figure 18. Directly Heated Cathode Structure of Type 866A

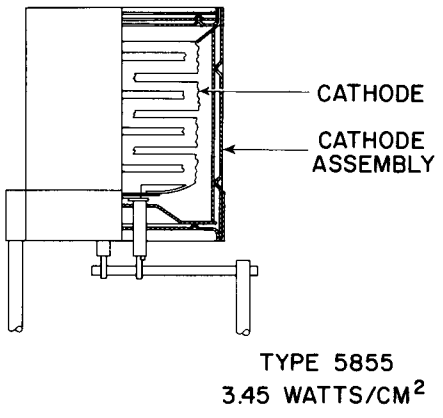


Figure 19. Directly Heated Cathode Structure of Type 5855

ing a total value of 0.008 seconds. For gas rectifiers, it has been customary to rate the peak current as four times the average current to cover the case of 180-degree conduction from a sine-wave source which yields a peak current of π times the average current. For industrial applications, where polyphase rectification and motor control are required, the conduction angle decreases and the ratio of peak current to average current rises. The design objective set up for thyratrons by the industry's standards organization (JEDEC) is 12.8 as the ratio of peak cathode current to average cathode current. Experience indicates that sufficient cathode area should be provided to avoid exceeding a peak current density of 2.0 amperes per square centimeter.

Anode Voltage. The figures for current densities given above are generally applicable to low-voltage tubes (below 2000 volts peak). Experience has shown that, for higher voltages, the current densities must be reduced to obtain the same life. The reasons for this interdependence are:

1. During the early parts of the inverse cycle of voltage, ions remain from the forward conduction cycle. These ions are accelerated toward the anode by the high negative voltage. Upon impact, foreign gases are released which poison the cathode.
2. If sufficient ions remain during the early parts of the inverse cycle of voltage, arcback may result.
3. The cathode current is thus limited, but not by the cathode. In the case of thyratrons, ions may be

accelerated by the high voltage during forward breakdown, and cause cathode bombardment and destruction.

Thermal Efficiency. After the cathode area required has been found, the power required to maintain this area at the optimum temperature of 1100 K must be determined. As described previously, the thermal efficiency of the cathode depends upon the amount of convoluting and the amount of heat shielding. The thermal efficiency is not subject to ready calculation. By comparison of a proposed structure with the samples given previously under Cathode Efficiency, one may obtain any value of thermal efficiency from 7 to 1 watt/cm^2. Note, when expressed in this fashion, a low figure means a high thermal efficiency.

Heat Shielding. The use of a large amount of heat shielding raises the thermal efficiency but incurs the disadvantage of a long heating time. Some highly shielded cathodes require as much as 5 minutes warm-up time which is objectionably long to the customer. More moderate amounts of heat shielding are used in modern gas-tube design. The heat shields should have low thermal emissivity on both inside and outside. The power transferred by radiation from the cathode through the heat shield to the atmosphere may be expressed as

$$q = A\,k(T_k^4 - T_0^4)\,v \text{ watts}$$

where A = cathode area
k = a constant (a function of shape)
T_k = cathode temperature
T_0 = outside temperature

and $$v = \frac{1}{1/e_1 + 1/e_2 + 1/e_3 - 1}$$

where e_1 = emissivity of cathode
e_2 = emissivity of inside heat shield
e_3 = emissivity of outside heat shield

Heater Design

If the indirectly heated cathode is chosen for a design, power must be transferred from a high-temperature heater to a lower temperature cathode. The above formula is useful to the extent that the power is radiated, but compact designs involve power conduction, usually through an insulating layer. Power transfer by both radiation and conduction is a complex problem not unique to gas tubes. The reader is referred to other portions of this book and to a paper by Cecil E. Haller.[16] This discussion hereafter will be confined to the directly heated cathode.

Voltage. The maximum heater voltage for a gas tube is limited to that value at which an end-to-end discharge may occur across the filament. The heater voltage, because it is ac, passes through a maximum of 1.41 times the rms heater voltage at 90 degrees phase angle. At this phase angle, one end of the filament is negative and capable of emitting electrons while the other end is positive and capable of receiving electrons. A discharge ensues providing the adjacent gas path becomes ionized. This process may occur without the presence of current to the anode. An end-to-end discharge will occur in conventional mercury tubes at heater voltages of 9 volts rms and higher; at high temperature, the resulting higher pressure lowers the threshold. This discharge may be observed on the oscillogram of heater current as a distortion of the sine wave near the peak.

By special design in which the heater is well-shielded, higher filament voltages (110 volts) may be used without showing end-to-end discharge.

Current. The combination of low heater voltage (2.5 volts) and high heater current is popular in modern gas tube design because large cathode currents may be permitted without appreciably affecting the heater temperature. A low instantaneous ratio of cathode currents to heater currents should be maintained at all times. In addition, the resulting low-resistance heater circuit permits the cathode current and the filament current to flow simultaneously without encountering high voltage drops. After the heater power and voltage have been determined the heater current may be calculated by using the equation I = P/E.

Resistance. The hot resistance of the heater may be calculated using Ohm's law R = E/I. The directly heated cathode is generally in the form of a ribbon which gives a maximum ratio of surface to volume. The following formula gives the length of the filament:

$$L = \left(\frac{RA^2}{4PB}\right)^{1/3}$$

where A = emitting area, both sides
R = hot resistance of filament
P = resistivity of filament material
B = ratio of width to thickness of ribbon.

Because the ribbon is made by rolling wire, the higher the value of B, the more expensive the ribbon. B has a maximum value of 40 for most materials. The width W can be determined from the formula W = A/2L.

In the manufacture of a filament similar to that of type 866A (Fig. 18) it is necessary to crimp the ribbon before winding it by threading it through a set of gears. Such crimping distorts the ribbon cross section so that the resistance may rise 5 to 10 per cent.

Experience has shown that to produce a directly heated filament from pure nickel for the standard low filament voltages of 2.5 to 5.0 volts requires such thin material that the strength is inadequate for conventional use. The resistivity of chemically pure nickel at the normal operating temperature of 837 C is 44 x 10^{-6} ohm-centimeter. The addition of cobalt to form a nickel-cobalt alloy raises the resistivity. An alloy of 40 per cent cobalt and 60 per cent nickel (known as RCA N97 material), which has a resistivity of 87 x 10^{-6} ohm-centimeters, is commonly used for directly-heated cathodes of gas tubes. For design purposes, the cold resistance of N97 material is 13.8 x 10^{-6} ohm-centimeter.

Gas Filling

Choice of Gas. Referring to Table II,[17,18] it may be observed that the alkali metals Cs, Rb, K, Na, Li,

offer a possible choice of gas filling since they exhibit some vapor pressure at a typical operating temperature of 100 C. The highest in this respect is mercury with cesium second. Mercury is used very successfully for the gas filling of electron tubes. Cesium has been tried on an experimental basis but no commercial tubes using it have been announced. All of the inert gases have been used for filling gas-discharge tubes or lamps. The inert gases with low ionization voltages yield the longest life in hot-cathode tubes due to the low energy of the ions that bombard the cathode. Argon was used extensively until xenon became commercially available. Because of their chemical activity or instability, other gases such as oxygen, nitrogen, carbon monoxide, and carbon dioxide are not used. Hydrogen is a notable exception. Research during World War II showed that hydrogen could be used provided exposed parts within the tube were degassed of oxygen. The amount of ion bombardment that a hot cathode can stand is described by the disintegration voltage.[19] The following voltages have been reported for the various gases:[20] mercury, -22; argon, -25; neon, -27; and hydrogen, -600.

Clean-Up. Another factor in determining the choice of gas for filling a gas tube is the gradual disappearance of the gas during life; this phenomenon is known as clean-up.[21] Gas ions, upon impact with the elements or the envelope, may remain on the surface or become imbedded within the body. The theory of gas clean-up is still incomplete, but its characteristics are well known. Gases such as nitrogen or oxygen clean up readily. For the noble gases, clean-up decreases with increasing weight. Xenon, because it is the heaviest, finds almost universal use. The low pressure (less than 100 microns) made necessary by the high inverse voltage ratings of gas tubes makes clean-up a severe problem. Clean-up may be reduced in a hot-cathode gas tube by reducing the anode-cathode spacing and by minimizing the surface area exposed to the discharge. In other words, shielding the discharge from exposure to wall surfaces of the envelope reduces clean-up. Operating conditions affect clean-up which increases with increasing inverse voltage, increasing forward current, and increasing frequency. During the time immediately following forward current conduction, ions remain within the envelope (see Fig. 20). The number of these ions is directly proportional to the electron current flowing prior to cessation of forward current (time T, Fig. 20). The energy with which these ions bombard exposed surfaces is proportional to the inverse voltage. This process, having a duration of 10 to 100 microseconds, occurs every cycle and, thus, clean-up is proportional to frequency. Average life expectancy, which is an inverse measure of clean-up, is given in Table III for typical xenon-filled rectifiers for 420 cycles per second.

Sputtering. Sputtering is the removal of particles of the incident surface by the bombarding ions. Generally, gas clean-up is quite severe when accompanied by sputtering. Studies of sputtering have shown there is a threshold of voltage, characteristic of each metal. These thresholds are shown in Table IV.[22] It is believed that the metals which sputter with difficulty will show the least gas clean-up.

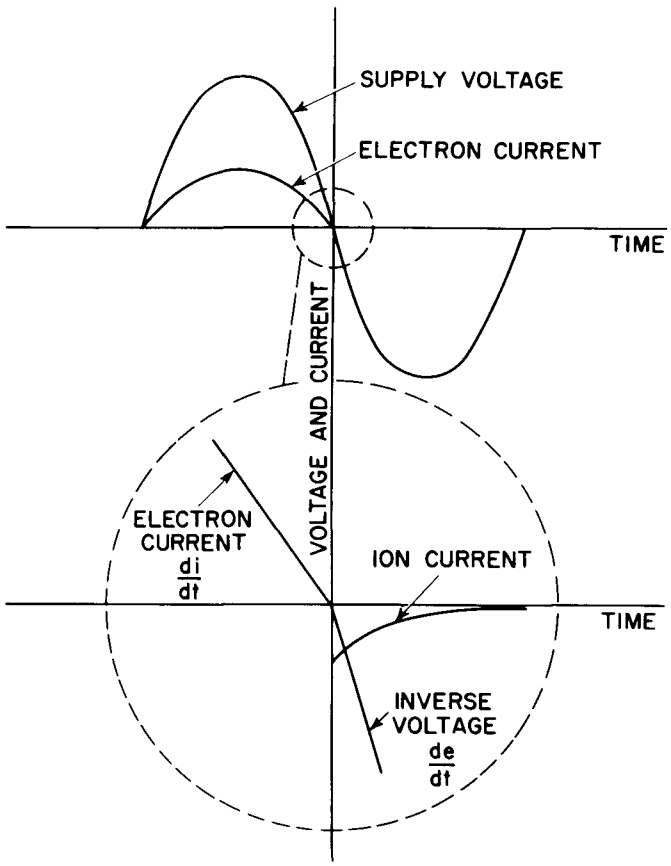


Figure 20. Commutation Factor

TABLE III. Average Life Expectancy

Type	Design	Inverse Voltage Kilovolts	Average Forward Current Amperes	Average Life Expectancy Hours
3B25	Unshielded	4.5 Kv	0.5	5000
3B28	Shielded	10.0	0.25	2000

For anode material, gas tubes employ copper, iron with zirconium spray, nickel, molybdenum, and tantalum. Copper, although subject to sputtering, is used due to its high thermal conductivity.

Poisoning. In addition to sputtering, another effect may occur when ions bombard a surface in a gas tube. Foreign gases which will poison the cathode may be ejected. An ion current of 1 milliampere bombarding a previously degassed nickel surface at 250 volts acceleration will poison a 1-square-centimeter cathode surface in 2-1/2 hours. This poisoning effect has been demonstrated with the type 2050 thyratron.[23]

Circuit Cushioning. Circuit cushioning is used to minimize the harmful effects of ion bombardment. In Fig. 20, the inverse voltage reaches an appreciable

TABLE IV.
Sputtering Threshold Voltages for Metals[22]

Metal	Volts
Columbium	99.8
Tantalum	97.1
Aluminum	90.8
Hafnium	90.7
Tungsten	89.5
Yttrium	89.0
Titanium	85.5
Molybdenum	76.5
Zirconium	72.8
Cobalt	70.3
Nickel	67.6
Iron	66.2
Rhodium	64.5
Chromium	55.7
Platinum	54.7
Copper	52.5
Palladium	42.0
Gold	36.2
Silver	33.4
Platinum	19.9

value before the ion current has decayed to a negligible value. The use of a combination of resistor and capacitor in series across a gas rectifier, will reduce the rate of rise of inverse voltage. Such a circuit is called a cushioning, or snubber, circuit.[24] The severity of ion bombardment is described partially by the commutation factor which is defined as the rate of fall of electron current di/dt (Fig. 20) multiplied by the rate of rise of inverse voltage de/dt. Many commercial tube types have been rated to withstand some maximum commutation factor. The use of these tubes must be restricted to circuits having commutation factors[25] within these limits.

The commutation factor depends upon tube design. The 5685 (6.4 amperes and 770 volts) with its open structure will stand only 0.66 volt-amperes per microsecond squared, whereas the close-spaced 6807 (6.4 amperes and 1500 volts) will take 130 volt-amperes per microsecond squared. To date commutation factor rating has been applied only to triodes, but it is equally applicable to diodes. To design for a high commutation factor requires that the following principles be incorporated:

(a) Use close spacing between anode and cathode or grid.

(b) Enclose discharge to restrict bombarding area.

(c) Degas bombarding areas well by r-f or by ion bombardment while on the pumps.

(d) For bombarded areas, use materials which resist sputtering.

(e) Use gas which does not clean up readily.

Reservoir. A reservoir may be used to overcome gas clean-up. The mercury tube has a natural reservoir in the drop of liquid mercury with which the tube is dosed. There is clean-up in a mercury tube but it is never observed because the liquid is capable of providing many fillings of gas. For example, in type 816 only 0.01 milligram of mercury is required to fill the tube to 25 microns pressure (60 C), but the average liquid[26] mercury content is 500 milligrams. The absorption and desorption characteristics of hydrogen by various metals permit the design of a reservoir for the hydrogen-filled tube. Although hydrogen reservoirs have been described using various metals such as zirconium, titanium, tantalum, cesium, and lanthanum, titanium has become the most popular. One gram of titanium can be loaded with 300 liter-millimeters of hydrogen which, if it all were available, could replace the gas in a liter-volume tube 300 times to a pressure of 1 millimeter; (0.6 millimeter is typical pressure in a hydrogen-filled tube). In practice, the reservoir operates as a ballast to maintain the pressure at the desired value by absorbing or desorbing as necessary. The equilibrium value is determined by the temperature of the reservoir. (A typical temperature is 400 C.)[27]

Dosing. Several methods are available for dosing or filling the tube with the chosen gas. If the filling is done in the gaseous state at room temperature, the source may be spectroscopically pure gas commercially available in flasks of 1/4 to 1 liter at atmospheric pressure. Two stopcocks may be employed in series to dose quantities into the system. On machine exhaust, a pressure regulator reduces the pressure to about 100 millimeters; a rotating valve permits further expansion to give a dosing pressure of about 100 microns.

On trolley exhaust, a porous valve (Fig. 21) may be used for fine control of the pressure. In the position shown in the illustration, no gas leaves the supply because the mercury closes up the pores. When the upper plug is brought in contact with the lower plug, mercury is squeezed out from between the two plugs and gas passes from the high pressure source slowly into the system. The rate can be held as low as 10 liter-microns per second. For hydrogen dosing, the palladium needle method is accurate and safe. In this method (see Fig. 22), forming gas (10% H_2 in N_2) is passed around the palladium needle which, when cold, is quite impervious to all gases. Application of power to the tungsten filament raises the temperature of the palladium needle which then becomes pervious only to hydrogen. In this manner, very pure hydrogen is admitted to the system at a controllable rate.

Several methods are in use for dosing mercury into tubes. In high-production low-cost tubes, the mercury is dosed through the exhaust tubulation by compressed air before evacuation. A surplus of mercury is dosed in this manner to cover the loss during bakeout and radio-frequency treatment. More accurate control over the final dosage is the hand-in-the-pellet method. In this method, mercury is contained in a glass or metal pellet which is attached to the mount before main sealing. In the later steps of the exhaust cycle, the mercury is released by cracking or exploding the pellet with radio-frequency current. European practice includes

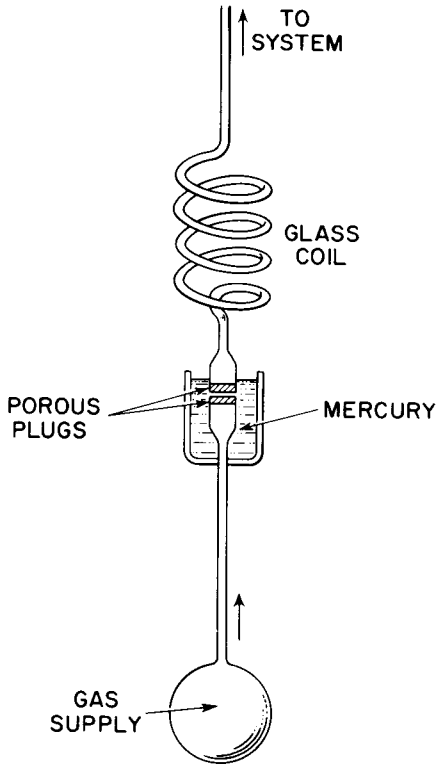


Figure 21. Porous Plug Valve

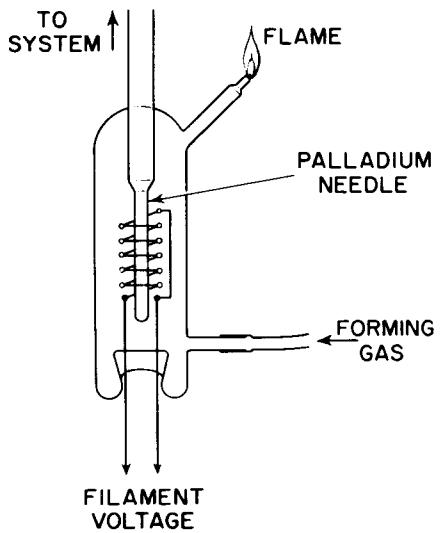


Figure 22. Palladium Needle Doser

a further refinement: a pellet containing mercuric oxide, zirconium, and iron is flashed at 500 C to decompose the oxide and release the mercury. The oxygen is gettered by the zirconium to avoid cathode poisoning.

Pressure. Paschen's Law (see Fig. 9) is a determining factor of the pressure to which a tube is dosed. To obtain a high operating voltage, the pressure must be above or below the minimum abscissa of Paschen's Law. Tubes dosed above the minimum exhibit a long deionization time which limits their usefulness even at 60 cycle-per-second operating frequency. Hence, practically all hot-cathode gas tubes are dosed below the minimum of Paschen's Law. For example, to raise the arc-back voltage of a xenon-filled tube to 1000 volts, the pressure-distance product must be less than 0.5 millimeter Hg-centimeter. Maximum long path between anode and cathode may be 2 centimeters which means the maximum dosing pressure is 0.25 millimeters or 250 microns.

In the case of mercury there is no control over the dosing pressure, but the operating pressure may be controlled by controlling the temperature of the liquid. The coolest surface inside the tube will condense the vapor to a liquid and will therefore be the critical temperature. Fig. 23 shows the relationship between the vapor pressure and liquid temperature of mercury.

Anode Design

The power input to a gas tube is the product of the tube drop and the cathode current. The electrons entering the discharge at the cathode have a very low probability of reaching the anode without a collision. In Fig. 5 it is shown that an electron accelerated to 10 volts in mercury vapor will have about 60 collisions in 1 centimeter. The electrons which are collected by the anode have been slowed down by collisions or have originated in the discharge. Thus the kinetic energy of the electrons collected by the anode is less than 10 volts. The average anode heating is about 5 watts per ampere,[28] depending upon the anode material and area relative to other surfaces exposed to the discharge. The remaining power is removed from the discharge by

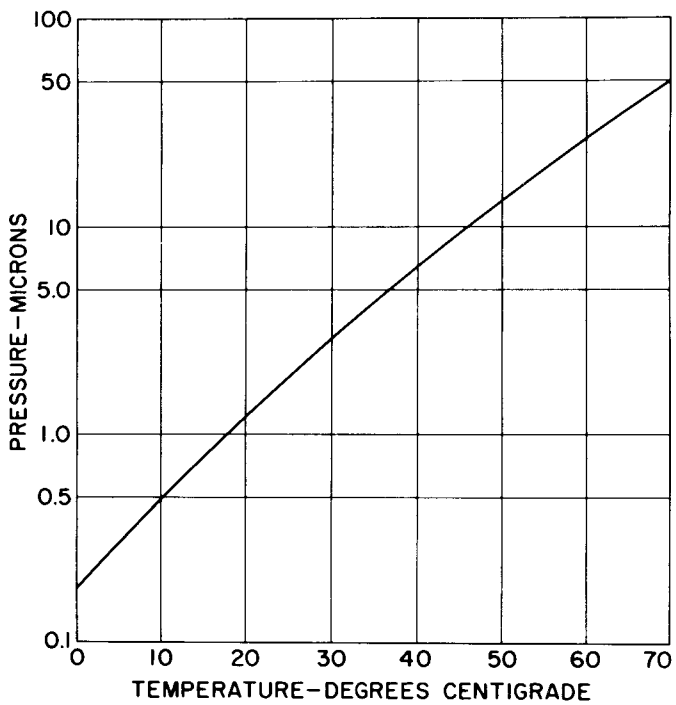


Figure 23. Vapor Pressure of Mercury

the ions and by radiation. Various materials are used for the anodes of gas tubes depending upon the application of the tube. For a rectifier where high thermal emission and low secondary emission caused by ion bombardment are important, carbonized nickel is a favorite material. Carbonized nickel which has thermal emissivity of 0.7 and operates at 300 C will radiate 0.37 watts per square centimeter. If the anode heating figure above is used the calculated anode area is 13.5 square centimeters per ampere. For carbon anodes, where the operating temperature may be considerably higher, one square centimeter per ampere is permissible. Tantalum and molybdenum are used for anode material of gas tubes because these materials resist the effects of ion bombardment on the inverse cycle (see Table IV) and have, as a result, low clean-up. Tantalum also has the advantage of gettering foreign gases when operated at about 1000 C. Zirconium is used as a coating for other metals for anodes. Zirconium hydride is mixed with iron oxide and is applied by spraying. Gettering action begins at 400 C. In one gas tube, copper has been used as an anode in order to readily conduct the heat to a radiating surface and to avoid hot spots. Other materials are described in reference 29.

Tube Drop

Measurement. If a direct current is passed through a gas tube, the voltage drop will show a constant or decreasing value as the current is increased. This effect on voltage drop results from the increase in cathode temperature caused by the increasing power injected into the discharge. Such is not the case for intermittent or periodic pulses of cathode current. Short duration pulses, i.e., one-half sine pulse from a 60 cycle-per-second source injected once per second, do not alter the cathode temperature and, therefore, give a true picture of the effect of high current on tube drop.

Effect of Pressure. The voltage-current characteristic of a gas discharge over many orders of magnitude of current is shown in Fig. 10. Tube drop is, by definition, the voltage across the tube, anode-to-cathode, subsequent to breakdown. Fig. 24 illustrates the manner in which tube drop varies with anode current. If the cathode is well shielded, as in the FG-166, there is a slight decrease in tube drop as current is increased from zero. After the minimum, there is a linear rise to the rated maximum peak current of the tube. In the case of an open cathode structure with little shielding (such as that of the 866A), the tube drop rises continuously from zero voltage without a minimum. The transition from the space-charge-limited current to the gas-discharge current is imperceptible. Note that at low current the tube drop is less than the ionization voltage for mercury (10.4 volts). Since the tube drop in this case is measured to the center tap, we must add half the peak filament voltage (1.75 volts) to the values given to get the peak circuit voltage. For the higher condensed mercury temperatures, the circuit voltage is still less than the ionization voltage. The explanation may be found in Fig. 11, which shows that the spacial distribution may exhibit a voltage difference greater than the circuit (cathode-to-anode) voltage.

Note, in Fig. 24, that the tube drop falls with increasing condensed mercury temperature. This fall results

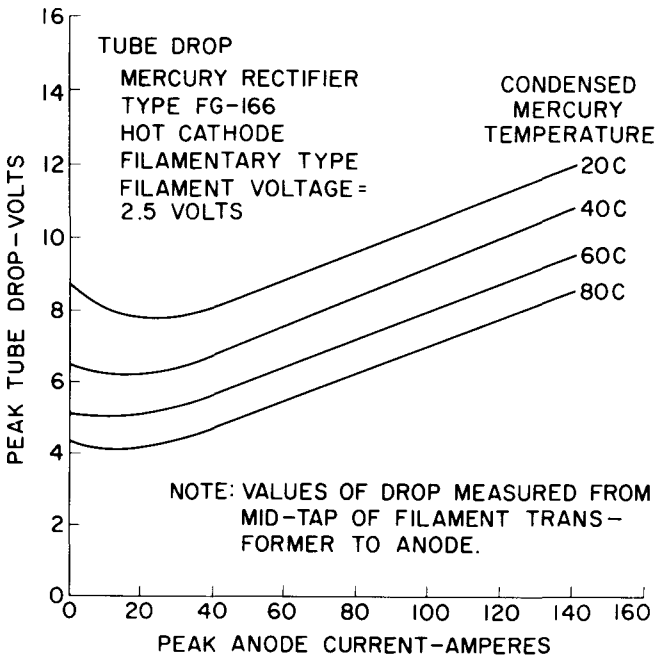


Figure 24. Tube Drop vs. Condensed Mercury Temperature

from rising mercury pressure which is accompanied by increasing probability of collision and greater ion density. A corollary to this relationship is the effect of decreasing pressure caused by clean-up in an inert gas tube. The rising tube drop increases the velocity of the positive ions. The high-speed ions may imbed themselves in the shield and cathode and thus accelerate the clean-up process. Furthermore, the rising tube drop will eventually reach the destructive sparking stage described later.

Effect of Filament Voltage. Fig. 25 shows how the slope of tube drop vs. current varies with the filament voltage (cathode temperature). The slope is an indicator of the quality of emission. For different cathodes operating at the same temperature, the one having the least slope has the best emission. During life the slope may fall slightly for about 100 hours, but thereafter will increase to the end of life. Fig. 26 shows how the tube drop increases as life progresses.

Sparking. For a given gas diode, if the pulsed cathode current is increased, the phenomenon of sparking will eventually occur. The voltage-drop trace on an oscilloscope will show a sudden discontinuity (Fig. 27) where the voltage drop will change from about 25 volts (for a xenon tube) to about 5 volts characteristic of the barium arc. Simultaneously, one or more incandescent particles can be seen in the tube flying from the cathode. Prolonged sparking will damage the cathode since a small spot of emitting area is destroyed at each spark. It has been found[30] that decreasing the pulse length permits higher peak currents to be drawn without sparking. Therefore, it has been concluded that the sparking results from a definite amount of energy dissipated in the cathode coating. Studies have shown that sparking

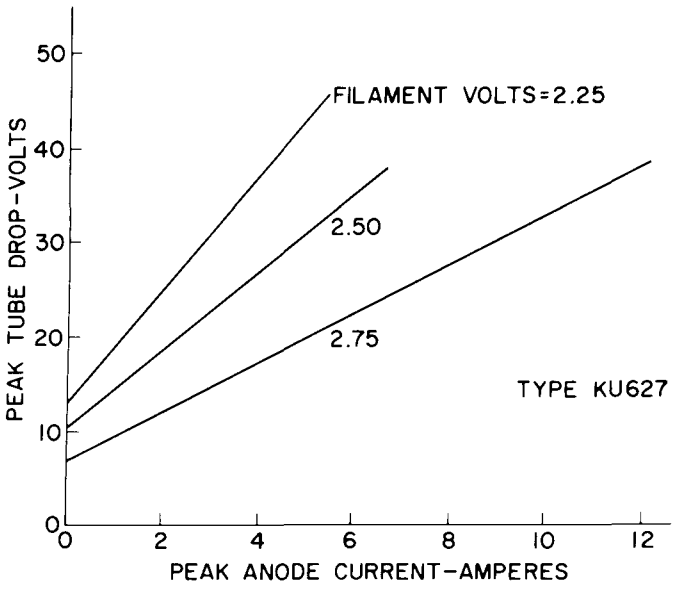


Figure 25. Tube Drop vs. Filament Voltage

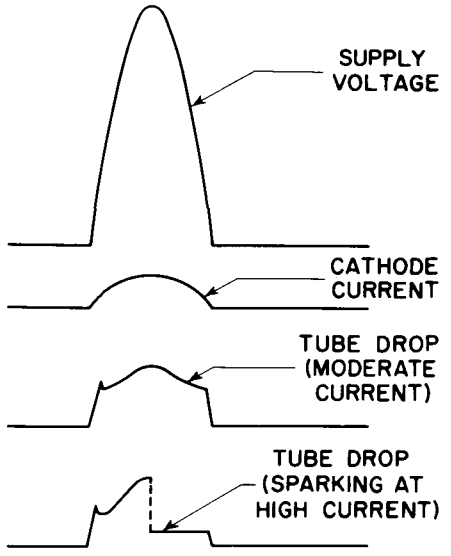


Figure 27. Sparking

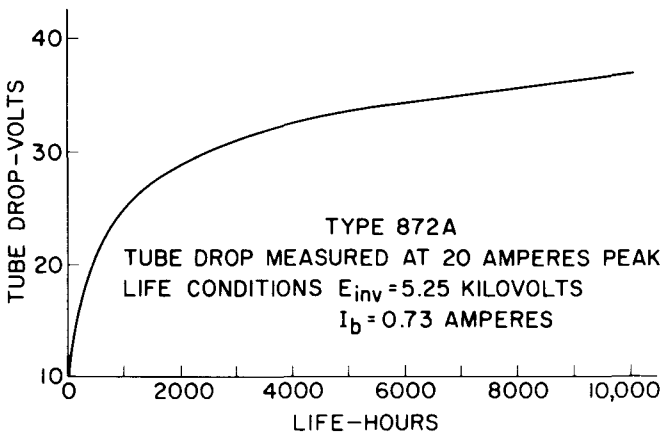


Figure 26. Tube Drop vs. Life

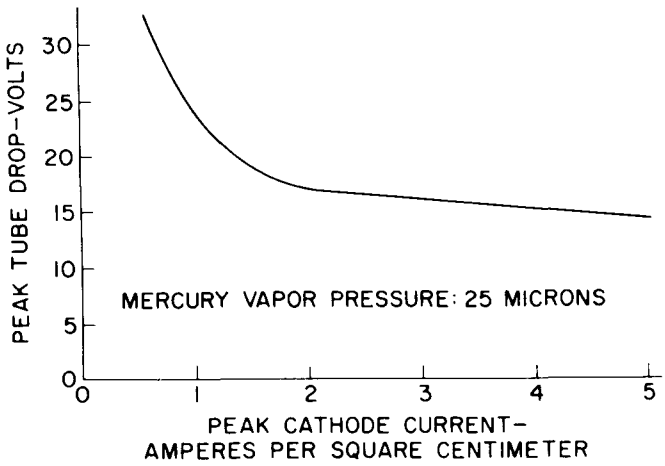


Figure 28. Sparking Limit Curve

tube drop vs. sparking current density is independent of cathode quality and filament current. Fig. 28 shows such a curve for mercury vapor at a pressure of 25 microns.

Ion Starvation. If the cathode in a gas diode is well shielded so that the discharge must traverse a hole of small cross section and appreciable length, another limitation on the maximum currents is encountered. The trace of voltage drop will show a discontinuity; the trace will rise to about the supply voltage and the current will cease for a short interval. Subsequently, the discharge will re-establish itself and oscillation may ensue. This phenomenon has been identified as ion starvation. The conducting path described has become starved of conducting ions. All of the gas molecules become ionized and the flow of ions to the cathode region depletes the conducting path of positive space neutralizers. The negative space charge of the electrons rises to that characteristic of a vacuum and causes the high voltage drop. Neutral particles rush in to re-establish the arc. The sharp discontinuities play havoc with the circuitry. Ion starvation can be avoided by designing for adequate conducting paths and using sufficiently high gas pressure during operation. Fig. 29 shows results of an investigation of ion starvation on mercury tubes.[31]

Fault Current. Gas tubes are rated for average current, peak current, and fault current. The average current is largely determined by the dissipation and hence by the size of the anode and of the bulb. The peak current is primarily a function of cathode design. The fault current is the current which may be permitted perhaps a dozen times during the tube life as a result of a fault in the circuitry. Both sparking and ion starvation must be considered in the establishment of a fault-current rating. In addition, the leads must be capable of carrying the fault current.

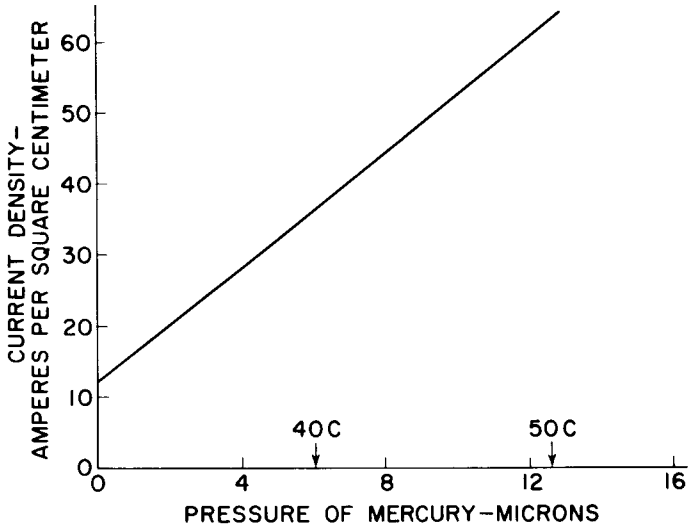


Figure 29. Ion Starvation

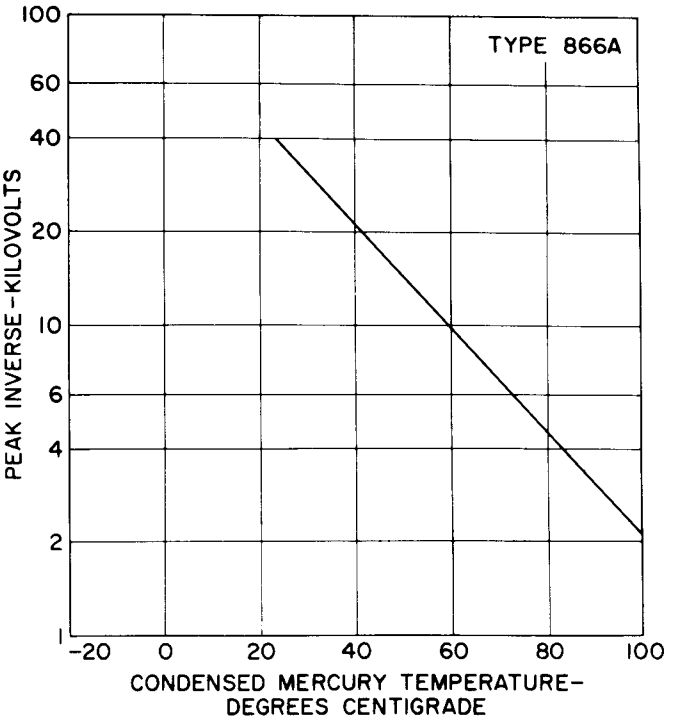


Figure 30. Arcback vs. Temperature

Arcback

Pressure. As described earlier, the breakdown of a gas between two electrodes is governed by Paschen's Law. Breakdown in the forward direction is at a low voltage because the externally heated cathode emits electrons without the aid of ion bombardment. In the reverse direction, breakdown is high because the normally cold anode will emit electrons only when bombarded by ions. Because the majority of gas rectifiers are filled to a low pressure, operation is on the lower side of the minimum of Paschen's Law.

The curve of Fig. 30 shows how the arcback voltage varies with condensed mercury temperature for the type 866A tube. Because pressure is a function of condensed mercury temperature, this curve represents the portion of Paschen's Law below the minimum. After equilibrium has been established, mercury vapor condenses on the coolest surface of the tube, and evaporation from this region determines the pressure. It is evident from Fig. 30 that high arcback voltage requires low condensed-mercury temperature. To obtain this condition, the designer must provide sufficient space and heat shields between the mercury-condensing zone and the heat producing elements within the tube. In the type 575A, a metal disk is inserted between the cathode and the mercury-condensing zone located just above the base. In other types, such as the FG-166, the mercury condensing zone is an appendage tubing fitted with radiating fins.

The pressure in inert-gas-filled tubes is not subject to the user's control. In such tubes, the designer must select a pressure which is not so high as to cause arcback and not so low as to clean up too rapidly. A practical aspect of the arcback problem is that tubes may arc back contrary to Paschen's Law for the known mercury pressure. The reason may be the presence of a foreign gas having a Paschen's Law different from that for mercury. Limited operation at high voltage will clean up the foreign gas. Such operation is known as seasoning or high-voltage aging.

If not retired for preventive maintenance, all gas rectifiers fail by arcback. Verification of this statement may be found in the arcback process just described. Foreign gas, evolved from the rising tube drop accompanying emission failure, has an unfavorable Paschen's Law which results in arcback at lower than rated voltage. Reduction of dc output voltage due to the emission failure is small in high-voltage circuits. Another cause of arcback in mercury tubes is localized high pressure. A deposit of mercury on the anode (as a result of shipment or condensation) may evaporate when the tube is operated. Temporarily, the coolest surface loses control over the mercury pressure until equilibrium is re-established. Arcbacks may occur in the interim.

Spacing. Paschen's Law further states that, to obtain high breakdown voltage, the distance between electrodes must be small. Close spacing between the anode and cathode is not difficult to obtain, but possible long paths between the same electrodes must be eliminated. Close-fitting bulb and metal shields are employed to this end.

Anode. High inverse breakdown requires that electron emission from the anode be minimized. Anodes made of materials having a high work function such as carbon (4.39 volts) or iron (4.36 volts)[32] will have low electron emission. Practically, the material chosen must also have a high work function when contaminated with oxygen or barium. Another method of reducing arcback is to shield the cathode so as to minimize evaporation of barium onto the anode. Likewise, avoidance of severe cathode activation schedules is desirable. Aging and sparking schedules reduce arcback partly by

cleaning the anode of barium and oxygen evaporation products.

Because thermionic emission increases with temperature, a cool anode will minimize arcbacks. Shielding from the source of heat and good thermal emissivity is helpful.

The shape of the anode is important. Curved surfaces do much to reduce the electric field and thereby to raise the arcback voltage.[33]

Patch Effect. Another cause of arcback is the patch effect[34] in which an insulated particle on the anode surface may pick up a charge. Although the charge may be small, the minute spacing produces a large electric field — large enough to initiate field emission; the result is arcback. High-voltage seasoning or "high potting" serves to burn off such particles.

Anode Lead Design. Good electrode design requires careful attention to the entrance and exit on the bulb for over-all success. The anode lead is particularly critical due to the enclosing high electric gradient. Paschen's Law, normally presented for linear voltage distribution, must be modified for this case. The possible arcback discharge paths originating at the anode may be interrupted by one or more glass pant legs surrounding the anode lead. Such glass pant legs (made of tubulation) must not touch the lead or each other. They also break up leakage paths. Most mercury rectifiers use one pant leg; the inert gas rectifier, 3B28, uses two pant legs, while the large inert gas rectifier, 4B32, employs three pant legs. In addition to pant legs, a reentrant structure is sometimes employed around the lead on the outside of the bulb. Fig. 31 shows a complex reentrant structure employed on the 5948 hydrogen thyratron. The path between the outer lead and the bulb, which may contain a plasma at near ground potential, is interrupted by a layer of ambient air.

THYRATRONS

Control Characteristic

Grid. The insertion of a third electrode, called the grid, between cathode and anode of a hot-cathode gas rectifier permits control over the forward breakdown voltage. A gas rectifier tube containing a grid is called a thyratron. Fig. 32 shows the relationship between the cathode-anode voltage E_a and the cathode-grid voltage E_g for a negative-grid thyratron. (Positive-grid thyratron design is adequately covered by United States Government research reports.)[35] With E_g at zero, the cathode-to-anode breakdown voltage at point a (Fig. 32) is the same as if the grid did not exist. As the grid is made more negative, the breakdown voltage (E_a) increases linearly. The portion of the curve between points a and c is the useful region. A small voltage on the grid (tens of volts) may hold off a large voltage on the anode (thousand of volts). The curve is shaded because it does not represent a function relationship between E_g and E_a but rather a dividing line between two regions. Starting below the line, there is no breakdown and no conduction of current between cathode and anode. Raising either quantity E_a or E_g in a positive direction takes a point on the graph across the curve whereupon breakdown suddenly occurs. Recrossing the curve in the reverse direction does not restore the nonconducting condition; the reason for this lack of restoration will become apparent as grid current is studied. In other words, the grid of a thyratron is able to start conduction but is not able to interrupt conduction or to vary the amount of current. The exceptional cases and constructions whereby the grid may perform the latter functions is beyond the scope of this treatment.[36, 37]

The slope of the useful portion of the control characteristic (a to c, Fig. 32) is defined as the control ratio, analagous to amplification factor in a vacuum triode. The analogy is restricted to the case where the parameter, anode current, is negligibly small.

The maximum positive anode voltage rating of a thyratron restricts operation to the linear portion of the

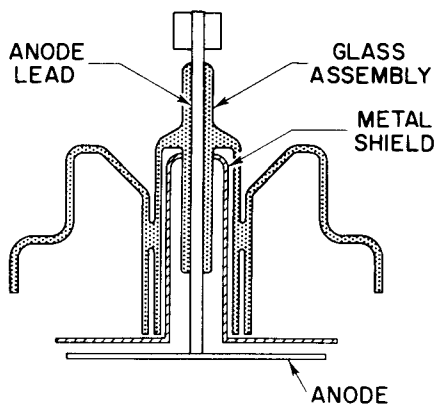


Figure 31. Reentrant Anode Lead Structure

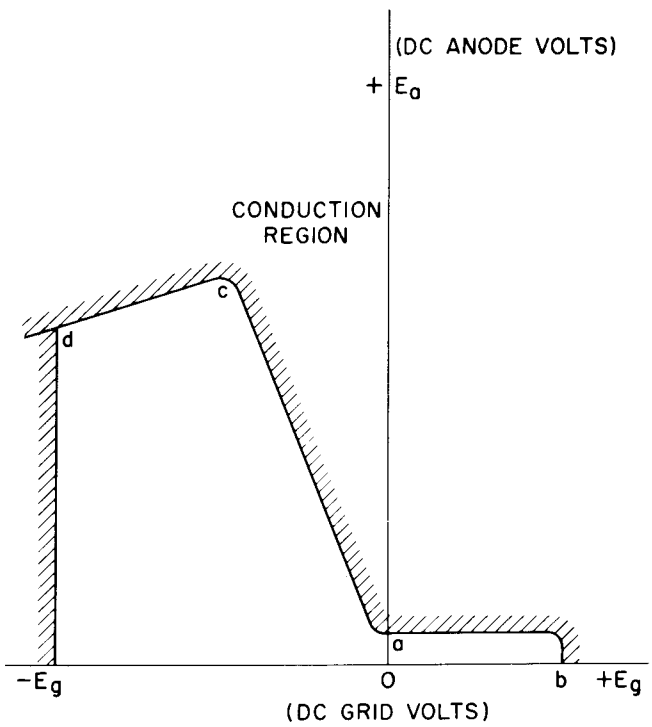


Figure 32. Thyratron Control Characteristic

control characteristic. However, if the anode voltage is increased beyond the rated maximum positive anode voltage, the grid loses control even though it be made very negative. This loss of control results from a breakdown occuring between the grid and anode in accordance with Paschen's Law; anode conduction normally follows. The downward slope (c to d) of Fig. 32 indicates that the grid-to-anode breakdown voltage is a constant value; this value is equal to the sum of the magnitudes of the two voltages referred to the cathode. The vertical line at point d is encountered when the negative grid voltage exceeds the grid-to-cathode breakdown voltage; again, this is just Paschen's Law working. In RCA Tube Division nomenclature this voltage is known as "grid glow volts." The magnitude of breakdown in the negative grid direction is several hundred volts. In the positive direction the control characteristic loses meaning when the grid-to-cathode path breaks down (point b) and the grid assumes the role of an anode. Here the magnitude is approximately equal to or less than the ionization potential (approximately 10 volts).

Theory. The voltage distribution in a typical thyratron before conduction is shown in Fig. 33. The space charge is negligible and, therefore, the equipotentials are the same as they would be in a vacuum tube. Fig. 34 shows the axial voltage distribution in the same thyratron before conduction. The hot cathode emits electrons of various initial velocities. All electrons with initial velocities less than that represented by the voltage minimum Vm are repelled and returned to the cathode. As the grid voltage is made less negative, a small portion of the electrons emitted are able to pass the voltage minimum and enter the grid-to-anode region. These electrons will collide with molecules producing ions and a small preconduction current. As the grid voltage is made less negative, the increasing number of ions begin to affect the voltage distribution (see dashed curve of Fig. 34). The positive ions have a greater effect than the electrons due to the slow speed of the ions. The rising equipotentials permit more electrons to pass and this creates more ions. Thus, the thyratron breaks down suddenly just as does a cold-cathode diode. Appendix D shows mathematically why the breakdown is sudden.

Effect of Pressure. In accord with Paschen's Law, the control characteristic of a thyratron is a function of pressure. At a given grid voltage, as may be observed in Fig. 35, the thyratron breaks down at a lower anode voltage as the pressure rises. Because the dosing pressure and the maximum spacings within the tube put the operating point of thyratrons on the low side of the

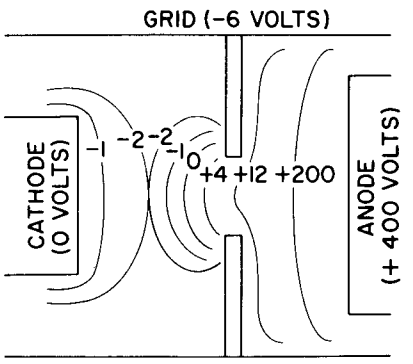


Figure 33. Equipotential Lines in Thyratron Before Conduction

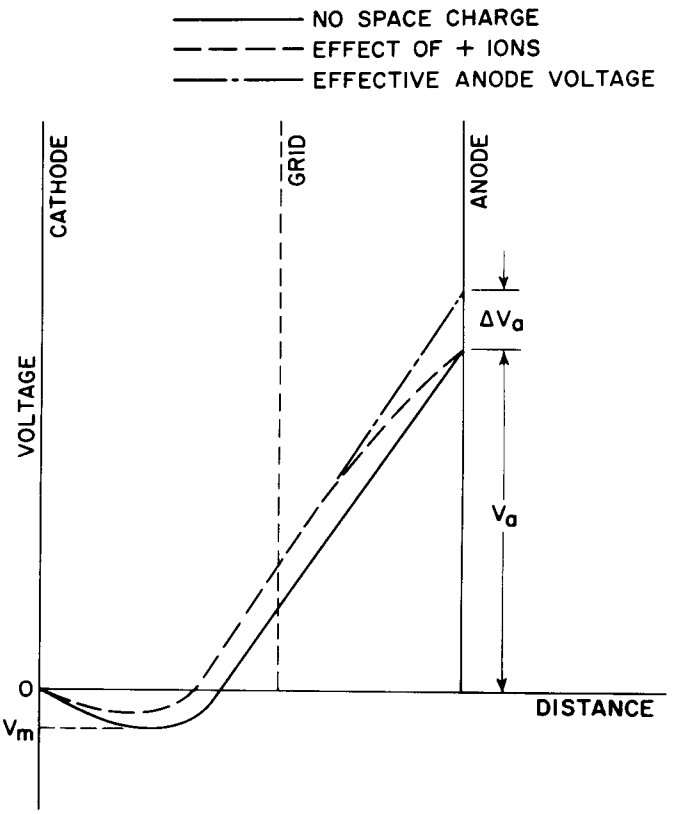


Figure 34. Axial Voltage Distribution in Thyratron Before Conduction

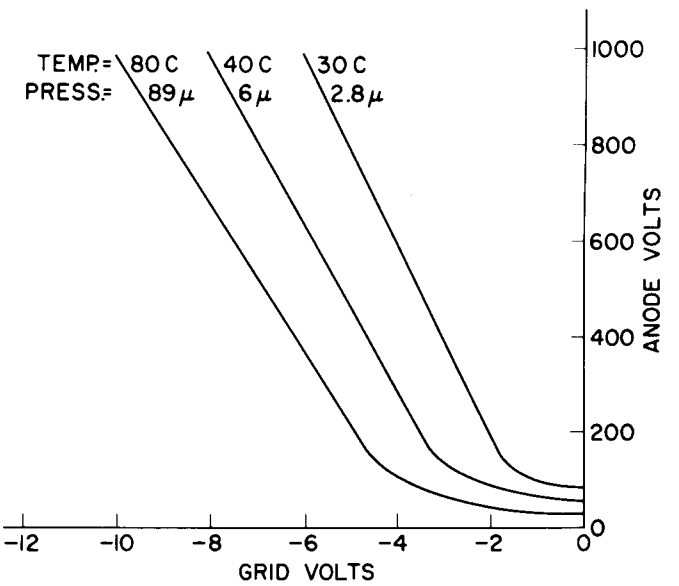


Figure 35. Control Characteristic of Thyratron at Different Pressures

minimum of Paschen's Law, it is logical to expect that thyratron breakdown would be similar to that of a gas diode.

Grid Current Prior to Conduction

Components. Fig. 36 shows the components which constitute grid current in a thyratron prior to conduction. Normally, the magnitude of these currents is in tenths of a microampere, but when the grid circuit resistance is in megohms, the grid currents become important. For Fig. 36, a positive dc voltage is applied to the anode. As the grid voltage is decreased from a negative value, the voltage minimum of Fig. 34 decreases in magnitude and some electrons surmount the hump. The grid collects some of these high energy electrons making up component curve 1 in Fig. 36. Many of these electrons pass through to the anode, because the grid hole (or holes) may be a large percentage of the grid cross section. The anode collects the remainder of these electrons making up curve I_a. As explained above, electrons reaching the grid-to-anode region are accelerated to ionizing velocity by the anode field and hence create ions. These ions flow in the opposite direction and are collected by the grid making up component 2 of Fig. 36; a leakage component 3 flows over the insulators on which the grid is mounted. Leakage is quite unpredictable; it generally decreases with grid voltage but may do so nonlinearly. Leakage may occur primarily to the anode, in which case it would be constant in Fig. 36. Furthermore, leakage varies with temperature. Curve 4, which represents electrons emitted by the grid, is shown as fixed since the electrons are generally attracted to the higher voltage of the anode. Likewise, grid emission is a function of temperature. Finally, capacitance current (curve 5) is shown as constant with grid voltage because most of it flows to the anode due to the higher voltage. In shield-grid thyratrons, the capacitance from grid to anode, and consequently the current from grid to anode, is reduced. The resulting current I_g is the algebraic sum of all the components.

Effect of Anode Voltage. Fig. 37 shows the grid-current versus grid-voltage characteristic of a small thyratron (type 2050) with anode voltage as a parameter. The straight horizontal portions probably are leakage and grid emission. No capacitance current flows because the anode voltage is dc. The appearance of the ion component is quite evident as grid voltage is decreased. The small circles represent the point where breakdown occurs. Of interest is the 25-volt curve showing how the electron current to the grid predominates at low grid voltage.

Effect of Pressure. Fig. 38 shows the effect of pressure on the preconduction grid current. As the pressure is increased, the probability of collision and the amount of ionization increase. Thus, the ion current appears at a more negative grid voltage. The breakdown point is shifted to the more negative grid voltage (see also Fig. 35).

Grid Current During Conduction

Effect of Anode Current. The grid current of a thyratron during conduction is of the order of milliamperes

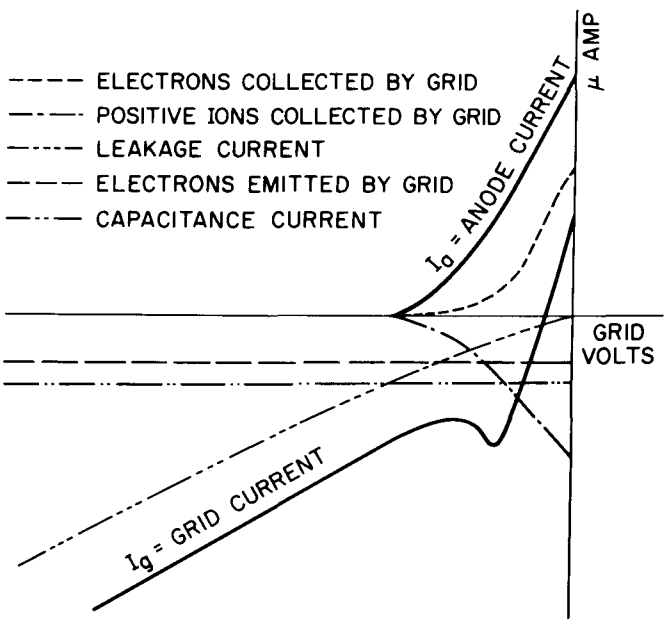


Figure 36. Components of Grid Current in Thyratron Prior to Conduction

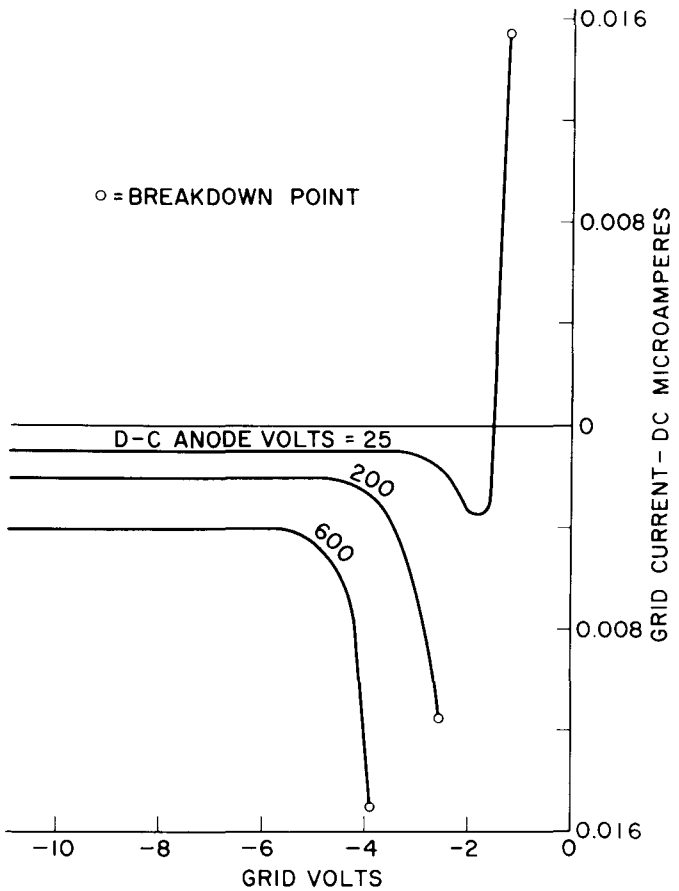


Figure 37. Grid Characteristic at Several Anode Voltages Prior to Conduction

– much greater than that measured before conduction. The grid is simply a probe in the discharge and is surrounded by a sheath across which ions flow to the grid from the plasma. The plasma has a random ion-current density determined by the anode current flowing. The flow of current is governed by the 3/2 power law. Fig. 39 shows how the grid current of a typical large thyratron (type 5563A) is saturated as the grid voltage is made more negative. As was explained earlier under "Probes," the sheath increases in thickness as the voltage increases, thus maintaining a constant current.

Effect of Pressure. Fig. 40 shows the grid-current versus grid-voltage characteristic with pressure as a parameter for a large thyratron (type 5563A). Although the random ion current of a plasma increases with pressure,[38] the voltage drop decreases (see Fig. 24) and the saturated current then decreases in accordance with the 3/2 power law.

Design

Shape. The grid of a thyratron may take many shapes. The conventional wire-wound grid of a vacuum triode functions poorly in a thyratron because the discharge may start at the edges of the coil. The grid in a thyratron must completely block all extraneous paths between cathode and anode and force the discharge through the intended aperture. The aperture may consist of holes or slots through a wall of negligible or appreciable thickness.

Spacing. Fig. 41 gives some data on spacing of the 5563A mercury thyratron. The grid has a single hole of considerable depth. In general, if either the grid-to-cathode or grid-to-anode spacing is increased, the control ratio (μ) increases. The reason for this behavior may be deduced from the equipotential plot of Fig. 33. The field of the anode reaches through the aperture to raise the equipotentials in the grid-to-cathode space. If the anode-to-grid spacing is increased, the anode voltage is not as effective and the control ratio increases. Similarly, if the cathode-to-grid spacing is increased, the anode voltage is less effective at the cathode and the control ratio increases. This relationship is valid for the type 5563A aperture which measures 0.563 inch in diameter by 0.520 inch deep.

Aperture Size. Fig. 42 gives a curve of the control ratio of a small shield-grid thyratron (type 2050) for variable slot width of the grid. As the slot width is increased, the anode field becomes more effective and the control ratio, therefore, decreases.

Material. The prime consideration in the choice of a grid material is grid emission, which is not unique with gas tubes. Several materials are being successfully used for gas tube grids. A large thyratron (type 7086) uses a copper grid for high thermal conductivity to reduce the grid temperature. The side exposed to the cathode is sprayed with aqua-dag to reduce grid emission. A high-voltage thyratron (type 5563A) uses carbonized nickel successfully. Gold-plated grids are

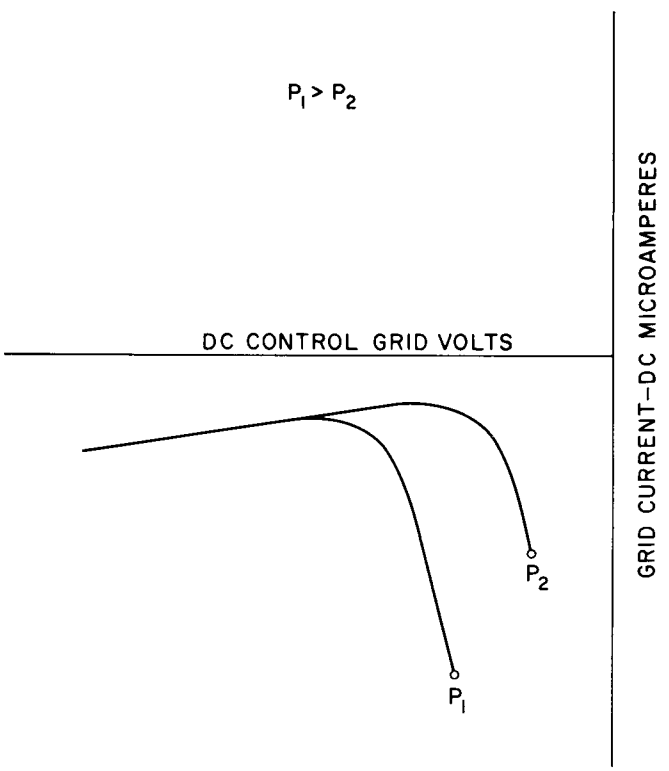


Figure 38. Grid Characteristic and Pressure Effect Prior to Conduction

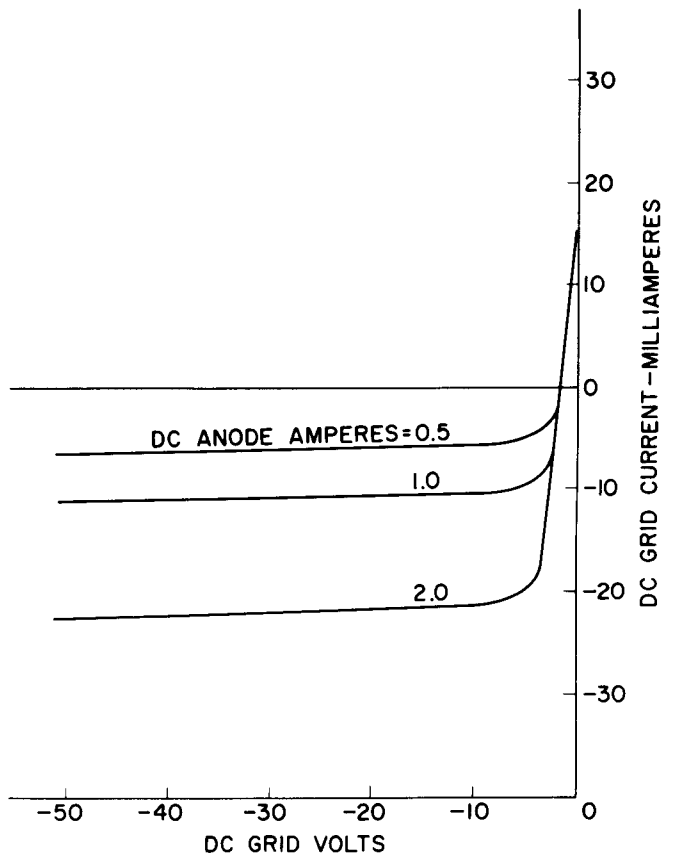


Figure 39. Grid Characteristic at Several Anode Currents During Conduction

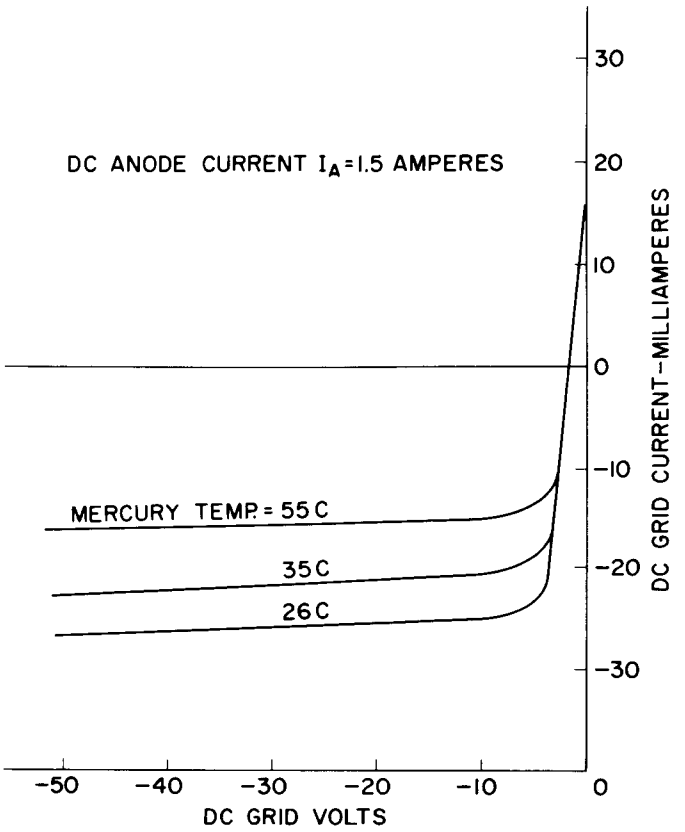


Figure 40. Grid Characteristic at Several Pressures During Conduction

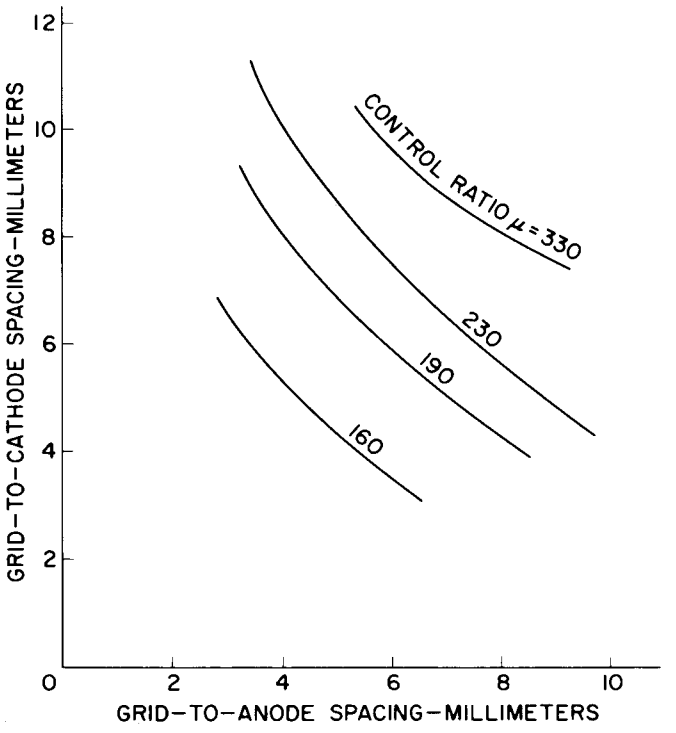


Figure 41. Control Ratio vs. Spacing

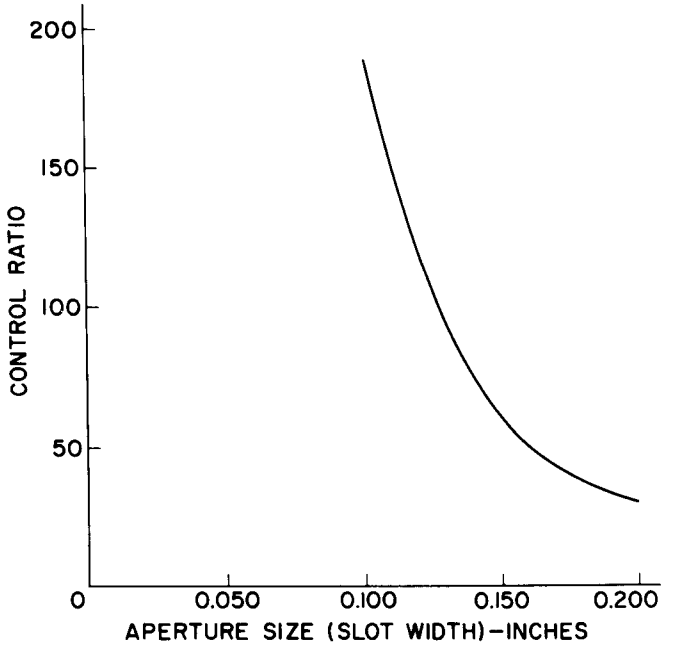


Figure 42. Control Ratio vs. Aperture (Slot) Size

used in small thyratrons, e.g., type 2050. Thyratrons merchandized by RCA, e.g., type C6J, use tungsten rods which have been heavily oxidized.

REFERENCES

1. Loeb, L. B., Kinetic Theory of Gases, 2nd Ed., p. 95., McGraw-Hill, 1934
2. Kennard, E. H., Kinetic Theory of Gases, McGraw-Hill, p. 149, 1938
3. Maxfield, F. A., and R. R. Benedict, Theory of Gaseous Conduction and Electronics, McGraw-Hill, p. 116, 1941
4. Parker, P., Electronics, E. Arnold Co., p. 581, 1950
5. Parker, P., op. cit.
6. Ibid., p. 591
7. Ibid., p. 593
8. Maxfield and Benedict, op. cit., p. 169
9. Druyvesteyn, M. J., and F. M. Penning, "The Mechanism of Electrical Discharges in Gases at Low Pressure," Rev. Mod. Phys., 12, 148, April 1940
10. Malter, L., E. O. Johnson, W. M. Webster, "Studies of Externally Heated Hot Cathode Arcs," RCA Review, 12, 417, September 1951
11. Applied Electronics, M. I. T. Staff, John Wiley, p. 120, 238, 1943
12. Wheatcroft, E. L. E., Gaseous Electrical Conductors Oxford Univ. Press, London, p. 187, 1938
13. Parker, P., op. cit., p. 1004
14. Parker, P., op. cit., p. 944
15. Maxfield and Benedict, op. cit., p. 118
16. Haller, E., "Filament and Heater Characteristics," Electronics, July 1944
17. Wheatcroft, op. cit., p. 30
18. Handbook of Chemistry and Physics, Chemical

Rubber Publishing Co., Cleveland, Ohio, 35th Edition
19. Hull, A. W., "Gas Filled Thermionic Tubes" A.I.E.E. Trans., 47, 753, 1928
20. Pulse Generators, M. I. T. Staff, McGraw-Hill, p. 336, 1948
21. Reddan, M. J., and G. F. Rouse, "Clean-up of Helium Gas in Arc Discharge," E.E. v. 71, 159, 1952
22. Code Q1
23. Code Q2
24. Edwards, D. V. and E. K. Smith, "Circuit Cushioning of Gas-Filled Grid-Control Rectifiers," A.I.E.E. Trans., 65, 640, Oct. 1946
25. Code Q3
26. Code Q4
27. Murphy, J. A., "Zirconium and Titanium Gas Reservoirs for Hydrogen Thyratrons," General Electric Co., Electronics Dept. Report 52E715, (8-12-52)
28. Wheatcroft, op. cit., p. 214
29. Spitzer, E. E., "Anode Materials for High Vacuum Tubes," E. E., Nov. 1935
30. Knowles, D. D., and J. W. McNall, "Sparking of Oxide Coated Cathodes in Mercury Vapor," J. A. P., 12, 149, Feb. 1941
31. Hull, A. W., and F. R. Elder, "The Cause of High Voltage Surges in Rectifier Circuits," J. A. P., 13, 372, June 1942
32. Reference Data for Radio Engineers, I. T. & T. Corp., 4th Ed., p. 43, 1956
33. RCA Bulletin, Types 6894, 6895
34. Cranberry, Louis, "Patch Theory of Arcback," J. A. P., 23, 518, May 1952
35. Goldberg, Seymour, Research Study on Hydrogen Thyratrons, Edgerton, Germeshausen, and Grier, Inc., Boston, Mass., Vol. I and II, 1956
36. Johnson, E. O., "Controllable Gas Diode," Electronics, 24, 107, May 1951
37. Knight, H. de B., "Hot Cathode Thyratrons," P.I.E.E. 96, 371, Sept. 1949
38. Parker, P., op. cit., p. 644
39. Hermann, G. and Wagener, S., Oxide-Coated Cathodes, Chapman & Hall, Vol. I, p. 78

APPENDIX A. BREAKDOWN

Let I_0 = number of electrons leaving negative electrode per second
α = ions or electrons produced per centimeter travel of electron
I_a = electron current reaching the positive electrode (anode)
x_a = distance negative to positive electrodes
I_p = primary electrons leaving cathode (exciting current)
I_+ = positive ion current reaching cathode
γ = secondary electrons released at cathode per incident ion

The electron current will build up as it crosses from the negative to the positive electrode. The increment per element of path dx is:

$$dI = \alpha I dx \tag{1}$$

Separating variables and integrating gives:

$$I_a = I_0 e^{\alpha x_a} \tag{2}$$

The positive ion current reaching the negative electrode is, by subtraction:

$$I_+ = I_0 (e^{\alpha x_a} - 1) \tag{3}$$

The current I_0 leaving the cathode consists of:

$$I_0 = I_p + \gamma I_+ \tag{4}$$

Combining Eqs. 2, 3, and 4:

$$I_a = \frac{I_p e^{\gamma x_a}}{1 - \gamma (e^{2x_a} - 1)} \tag{5}$$

The constants γ, α, and x_a may be given finite values such that the denominator becomes infinite; hence, mathematically, there can be a breakdown.

APPENDIX B. PASCHEN'S LAW

Eq. (5) of Appendix A becomes the breakdown equation when the current I_a is increased to a critical value cI_a. The variables remaining in Eq. (5) of Appendix A at breakdown are:

αx_a = the product of ionizing coefficient and distance between electrodes
γ = the secondary emission coefficient of the cathode

The ionizing coefficient α (see Fig. 5) is a function of the average number of collisions per centimeter $(1/\lambda e)$ (see Fig. 4) and the kinetic energy V of the colliding electron.

thus $\alpha = f_1 (\frac{1}{\lambda e}) f_2 (V)$ (1)

where λe is the mean free path of the electron

Substituting for V

$$\alpha = f_1 (\frac{1}{\lambda e}) f_2 (\frac{V_a \lambda e}{x_a}) \tag{2}$$

where V_a is the voltage between electrodes

Although λe varies with voltage, as Fig. 4 shows, above about 10 volts λe approaches the kinetic theory value and can be considered a constant. As the pressure is changed, λe varies inversely.

Substituting

$$\alpha = f_1(p) f_2 (\frac{V_a}{x a_p}) \tag{3}$$

Multiplying by X_a

$$\alpha X_a = f_3 (X_a P) f_2 (\frac{V_a}{X_a P}) \tag{4}$$

Thus one variable αX_a in the instability (Eq. (5) of Appendix A) above has been reduced to functions of $X_a P$ and V_a.

The secondary emission coefficient γ is a function only of the energy with which the ion strikes the cathode

$$\gamma = f_4 (\frac{\lambda i V_a}{X_a}) \tag{5}$$

where λi is the mean free path of ion.

Like λ_e, λ_i varies inversely with pressure

$$\gamma = f_5 (\frac{V_a}{X_a P}) \tag{6}$$

Thus the variables of the breakdown equation have been reduced to V_a and $X_a P$; the equation relating these two variables is Paschen's Law.

APPENDIX C. COLLISION OF ELECTRON AND ION

$$\frac{M_1 v_1^2}{2} = \frac{M_1 V_1^2}{2} + \frac{M_2 V_2^2}{2} \tag{1}$$

$$M_1 v_1 = M_1 V_1 + M_2 V_2 \tag{2}$$

where M_1 = mass of electron
M_2 = mass of ion
v_1 = velocity of electron before impact
V_1 = velocity of electron after impact
V_2 = velocity of ion after impact

Solving these two equations for the kinetic energy after impact of the electron gives:

$$W_1 = (1 - \frac{2M_2}{M_1 + M_2})^2 w_1$$

w_1 = kinetic energy of electron before impact
W_1 = kinetic energy of electron after impact

Because the mass of the electron M_1 is so small com-

pared to the mass of the ion M_2, the quantity in parenthesis is nearly unity and hence there is very little loss of kinetic energy by the electron in impact. For the electron to recombine, it must release all of its kinetic energy to the ion. The above relationship precludes this release and recombination is rare in the gas.

APPENDIX D. BREAKDOWN OF A THYRATRON

The solid line in Fig. 34 shows the axial voltage distribution of a thyratron without space charge. Of the electrons emitted by the cathode, only those having more kinetic energy than the potential energy of the voltage minimum pass through to the grid-anode region. Dividing both sides of Eq. (17) (An expression for the number of charged particles crossing unit area against a retarding voltage) by time converts the number to current:

$$I = I_s e^{\frac{qVm}{KT}} \tag{1}$$

where I_S is the total cathode emission
q is the charge and is negative.

The potential minimum V_m is a function ϕ, of V_g and V_a

$$V_m = \phi \left(V_g + \frac{V_a}{\mu}\right) \tag{2}$$

μ = control ratio.

As Fig. 34 shows, positive ion space charge raises the voltage distribution by an effective amount ΔV_a. Assume that the amount of rise ΔV_a is proportional to the anode current I_a.

then $$\Delta V_a = RI_a \tag{3}$$

where R is constant

Substituting in Eq. (1) gives:

$$I = I_s e^{\frac{q}{KT}\phi \left(V_g + \frac{V_a + RI_a}{\mu}\right)} \tag{4}$$

Taking the logarithm of both sides

$$\ln I = \ln I_s + \frac{q}{KT}\phi \left(V_g + \frac{V_a + RI_a}{\mu}\right) \tag{5}$$

Since I_S is a constant

$$\ln I = \phi \left(V_g + \frac{V_a + RI_a}{\mu}\right) \tag{6}$$

where ϕ is an unknown function

Now I is the electron current passing the voltage minimum. The anode current I_a is augmented by the ionization, but is a function of I.

Hence:

$$\ln I_a = f \left(V_g + \frac{V_a + RI_a}{\mu}\right) \tag{7}$$

Let the function f be a linear one

$$\ln I_a = K \left(V_g + \frac{V_a + RI_a}{\mu}\right) \tag{8}$$

K = constant

Rearranging

$$\ln I_a - K\frac{RI_a}{\mu} = K \left(V_g + \frac{V_a}{\mu}\right) \tag{9}$$

Taking the derivative of Eq. (9)

$$\frac{dI_a}{dV_g} = \frac{K}{\frac{1}{I_a} - \frac{KR}{\mu}} \tag{10}$$

In conclusion, $\frac{dI_a}{dV_g}$ becomes infinite when $1/I_a = -KR/\mu$; I_a is no longer limited by V_g and breakdown occurs.

Design and Processing of Oscillograph Tubes

E. M. Smith

Lancaster

Oscillograph tubes are cathode-ray display devices similar to the more familar television picture tubes. The important difference between the picture tube and the oscillograph tube is the application for which each is designed.

Oscillograph tubes are employed in many types of equipment. The cathode-ray oscilloscope is probably the most familiar. Oscillograph tubes are also used in flying-spot scanning, projection, TV-picture recording, and radar equipment.

COMPARISON OF MAGNETIC AND ELECTROSTATIC DEFLECTION

There are two basic types of oscillograph tubes: first, those which use electrostatic deflection; and second, those which are magnetically deflected. In electrostatic-deflection types, the electron beam is deflected by means of a voltage difference between two deflecting electrodes. A usual type uses two pairs of deflecting electrodes mounted so that the deflection produced by one pair is at right angles to the deflection produced by the other pair. The angle through which the beam is deflected by a pair of deflecting electrodes is directly proportional to the difference in potential between the electrodes, and inversly proportional to the ultor voltage.

The beam is usually deflected by the deflecting electrodes coupled to the outputs of voltage amplifiers which apply appropriate voltages to the electrodes. If the amplifiers have broad frequency response, the beam can be deflected over a wide range of sweep rates and at random intervals.

In magnetic-deflection types, the electron beam is deflected by a magnetic field at right angles to the beam path. This field is usually produced by a pair of coils mounted transverse to the neck of the tube. Unlike the arrangement of two pairs of deflecting electrodes for electrostatic deflection, it is possible and advantageous to place the second pair of coils on a magnetic-deflection tube so that the axis and fields of the two pairs of coils intersect. The angle through which the electron beam is deflected by a magnetic field is proportional to the current in the coils and is inversly proportional to the square root of the ultor voltage.

Because the deflection decreases in proportion to the accelerating voltage for electrostatic deflection, and as the square root of the accelerating voltage for magnetic deflection, the use of magnetic deflection becomes relatively more favorable as the ultor voltage is increased.

Because the relatively high inductive reactance of the magnetic-deflecting coils limits the frequency of the deflecting current which can be used, magnetic deflection is not feasible for displays requiring rapid changes in beam position. In most applications using magnetic deflection, the beam is deflected at regular intervals and at moderate rates, such as are used, for example, in television and in most radar displays.

Both electrostatic and magnetic deflection are accompanied by a deterioration of focus known as deflection defocusing. For high beam currents and wide deflection angles, deflection defocusing is five to ten times greater for electrostatic-deflection systems than for magnetic-deflection systems. For this reason, few electrostatic-deflection tubes have an over-all deflection angle greater than 30 degrees, whereas an over-all deflection angle of 110 degrees is common with magnetic-deflection tubes.

In applications involving a wide range of deflection frequencies or a rapid shifting of the beam from one area of the screen to another, electrostatic deflection is usually preferred. On the other hand, magnetic deflection offers advantages where the beam must be deflected over wide angles or operated at high ultor voltages. Magnetic deflection is preferred for tubes with display patterns as is the case in radar and television.

Figs. 1 and 2 show typical electron-gun designs for electrostatic-deflection and magnetic-deflection oscillograph tubes. Typical electrode voltages are indicated. An obvious difference between these two guns is the grid No. 2 voltage. In the electrostatic-deflection types, grid No. 2 is operated at the highest accelerating voltage, whereas the grid No. 2 of the magnetic types is operated at approximately 250 volts, and the highest voltage is applied to another electrode.

The use of a low-voltage grid No. 2 in the magnetic-deflection type of gun is important in that changes in the ultor voltage, caused for instance, by variations in beam current, will have relatively little effect on the control characteristic and on cutoff. Because magnetic-

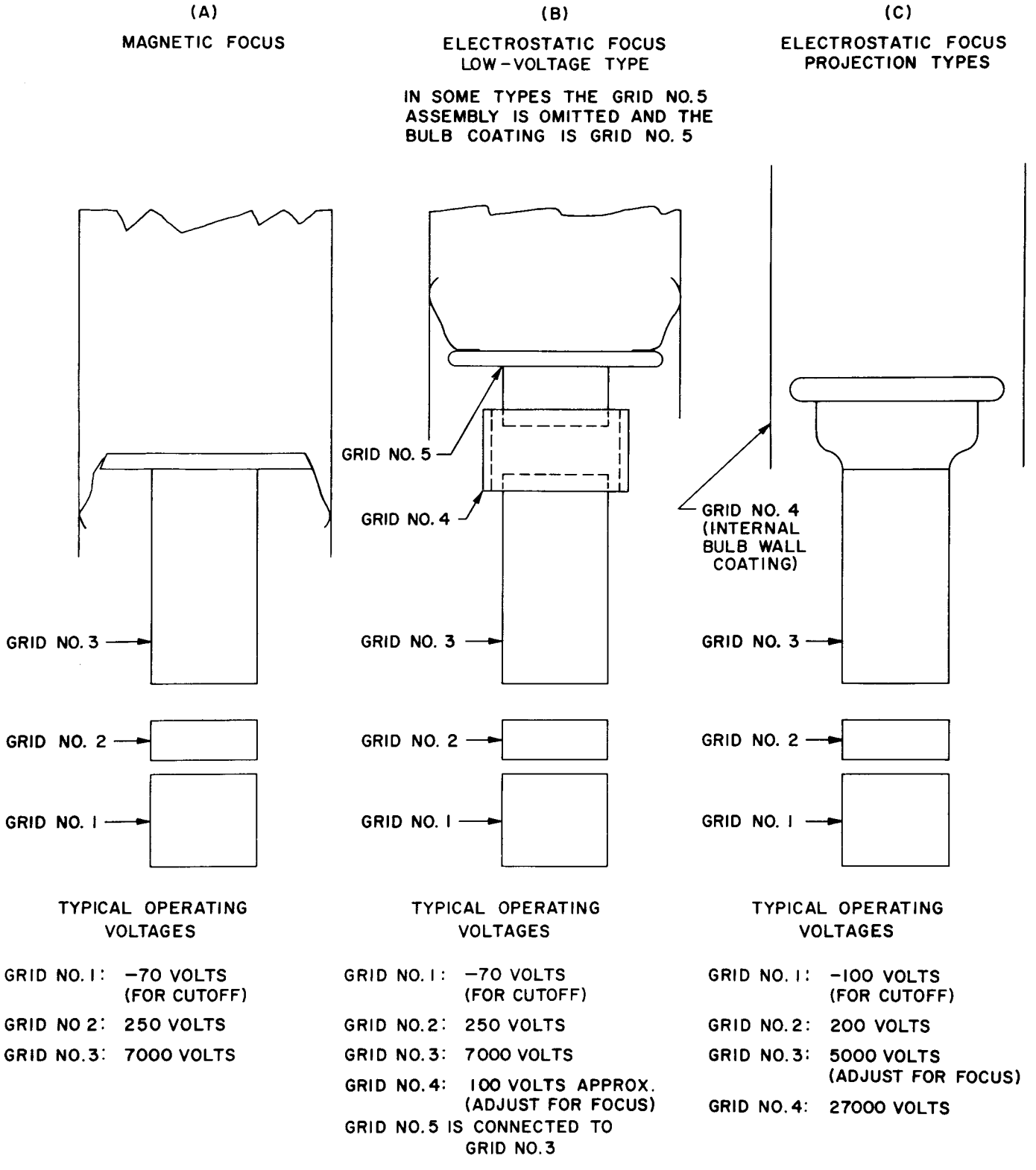
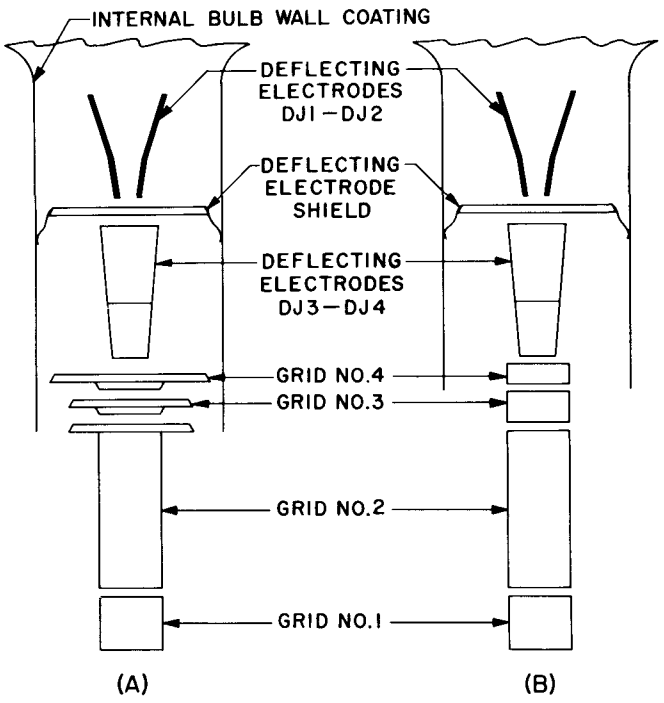

(A)
MAGNETIC FOCUS
(B)
ELECTROSTATIC FOCUS
LOW-VOLTAGE TYPE
IN SOME TYPES THE GRID NO. 5 ASSEMBLY IS OMITTED AND THE BULB COATING IS GRID NO. 5
(C)
ELECTROSTATIC FOCUS
PROJECTION TYPES
GRID NO. 5
GRID NO. 4
GRID NO. 4 (INTERNAL BULB WALL COATING)
GRID NO. 3
GRID NO. 2
GRID NO. 1
TYPICAL OPERATING VOLTAGES
GRID NO. 1: −70 VOLTS (FOR CUTOFF)
GRID NO 2: 250 VOLTS
GRID NO. 3: 7000 VOLTS
TYPICAL OPERATING VOLTAGES
GRID NO. 1: −70 VOLTS (FOR CUTOFF)
GRID NO. 2: 250 VOLTS
GRID NO. 3: 7000 VOLTS
GRID NO. 4: 100 VOLTS APPROX. (ADJUST FOR FOCUS)
GRID NO. 5 IS CONNECTED TO GRID NO. 3
TYPICAL OPERATING VOLTAGES
GRID NO. 1: −100 VOLTS (FOR CUTOFF)
GRID NO. 2: 200 VOLTS
GRID NO. 3: 5000 VOLTS (ADJUST FOR FOCUS)
GRID NO. 4: 27000 VOLTS

Figure 1. Typical Guns for Magnetic-Deflection Tubes

deflection types are employed where the display is produced by varying the beam current as the beam is swept across the screen, a constant control characteristic is important.

The electrostatic-deflection type of gun provides better resolution at the relatively low ultor voltage used in electrostatic deflection types. Also, for most applications where electrostatic deflection is required, grid No. 1 is not used for control other than as an on-off device.



TYPICAL OPERATING VOLTAGES, (A) AND (B)

GRID NO. 1: —50 VOLTS (FOR CUTOFF) MAX.
GRID NO. 2: 1500 VOLTS
GRID NO. 3: 380 VOLTS APPROX. (ADJUST FOR FOCUS)
GRID NO. 4, THE BULB WALL COATING, AND THE DEFLECTING ELECTRODE SHIELD ARE CONNECTED TO GRID NO. 2.
AVERAGE POTENTIAL OF EACH PAIR OF DEFLECTING ELECTRODES: 1500 VOLTS

Figure 2. Typical Guns for Electrostatic-Deflection Tubes

ELECTRON GUN DESIGN—ELECTROSTATIC-DEFLECTION TYPES

Refer to Fig. 2 and consider the electrostatic-deflection gun in more detail. The number of electrons which leave the unipotential, indirectly heated cathode and enter the lower grid No. 2 aperture is controlled by the voltage on grid No. 1. When a sufficiently negative voltage with respect to the cathode is applied, the electrons emitted from the cathode are turned back toward the cathode, thereby producing a space charge between the cathode and the grid. When a less negative voltage is applied to grid No. 1, the accelerating field of grid No. 2 can penetrate the grid No. 1 aperture and accelerate electrons from the space-charge region toward the screen. The electron optics system comprising the cathode, grid No. 1, and the grid No. 2 aperture nearer the cathode is quite complex, but is, nevertheless, important to the performance of the tube. This combination of three electrodes, known as the "triode" section of the gun, determines cutoff, beam current, drive characteristics, and, to a large extent, the resolution of the tube.

Between grid No. 1 and grid No. 2, the electrons which pass through the grid No. 1 aperture, cross over the beam axis to form a minimum cross-sectional area known as the cross-over. The cross-over is focused on the screen by the focusing lens to provide the minimum spot size. If a perfect focusing lens is assumed, the spot size in a given tube is directly proportional to the cross-over size.

Beyond the cross-over, the beam electrons diverge in the grid No. 2 cylinder. At the end of the grid No. 2 cylinder is an aperture which masks a part of the beam. The part of the beam which passes through the masking aperture is then converged by the focusing lens, which consists of grids No. 2, No. 3, and No. 4. Beyond the focusing lens, the beam is deflected first by the pair of deflecting electrodes nearer to the cathode (DJ3-DJ4) and then by the pair of deflecting electrodes farther from the cathode (DJ1-DJ2). The two pairs of electrodes are angularly displaced about the gun axis by 90 degrees. After the electron beam leaves the influence of the deflecting fields between the two pairs of deflecting electrodes, the beam follows a straight line (assuming no post-deflection acceleration) to the screen.

Under normal conditions, the focusing lens will be adjusted so that the beam focuses at the screen. The size of the spot produced by the electron gun or the resolution of the tube is determined by (1) the size of the cross-over, (2) magnification, (3) operating voltages, and (4) aberrations in the gun. For minimum cross-over size, the accelerating voltage (grid No. 2 voltage) should be as high as possible, and the grid drive (voltage above cutoff) to produce the required light output, should be as low as possible. Magnification of the gun is the ratio of the distance from the effective center of the focusing lens to the screen divided by the distance from the cross-over to the effective center of the focusing lens. Neglecting other factors, this ratio should be kept small to improve resolution. The general relation, spot size = cross-over size x magnification x (cross-over potential)$^{1/2}$ / (accelerating potential),$^{1/2}$ indicates the relative importance of the various parameters.

Aberrations in the gun occur in the triode section, in the focusing lens, and in the deflecting-electrode system. Aberrations can be reduced by limiting the size of the grid No. 1 aperture and the size of the masking aperture in grid No. 2; the diameter of the focusing lens should be made as large as possible. Aberrations and defocusing of the beam in the deflecting system can be reduced by limiting the beam diameter in that region. Aberrations in the other sections of the gun can be minimized by using clean, round apertures mounted so that the grid apertures and cylinders are exactly concentric.

The following tables may serve as a guide in determining a workable gun design. For a more complete study of these factors see Refs. 1 and 2 and in the article on Color Kinescope Design and Processing.

Table I lists the variables in gun design and indicates the effect on performance characteristics of an increase in the value of each variable. An increase in performance characteristic is indicated by +, a decrease by —. Table II presents the voltage vs. current relationships that are of importance in gun design.

Table I

Effect of Change in Variables on Performance Characteristics

Variable	Performance Characteristics Affected	Effect on Performance Characteristic of Increase of Variable
Length of grid No. 2 cylinder	Magnification -------	-
	Ratio: Beam current/ Cathode current -----	-
Top grid No. 2 aperture diameter	Ratio: Beam current/ Cathode current -----	+
	Beam diameter in focus lens and in the region of the deflecting electrodes -------	+
Diameter of grid No. 2 aperture closer to the cathode	Cutoff voltage -------	-
Grid No. 3 aperture diameter	Focus voltage -------	-
	Aberrations in focusing lens----------	-
Length of grid No. 3 cylinder	Focus voltage -------	+
Spacing between grid No. 2, grid No. 3, and grid No. 4	Focus voltage -------	+
	Maximum voltage rating for grid No. 3 -	+
Length of deflecting electrodes	Deflection factor-----	-
	Maximum deflection angle ----------	-
Spacing of the deflect- electrodes	Deflection factor-----	+
	Maximum deflection angle ----------	+
Grid No. 1 to grid No. 2 spacing	Cutoff ----------	-
	Maximum rated voltage----------	+
Grid No. 1 to cathode spacing	Cutoff ----------	-
Grid No. 1 aperture diameter	Cutoff ----------	+

Table II

Voltage-Current Relations

Deflection factor = $k_1 E_{c4}$
Cutoff voltage (CoE_{c1}) = $k_2 E_{c4}$
Focus voltage = $k_3 E_{c4}$
Maximum cathode current (zero bias) = $k_4 (CoE_{c1})^{3/2}$

INTERELECTRODE CAPACITANCE

The latest electrostatic-deflection oscillograph tubes are designed to have high deflection sensitivity and minimum interelectrode capacitances. Deflection sensitivity is increased (deflection factor reduced) by the use of close-spaced deflecting electrodes, small deflection angles, and long deflecting electrodes. The optimum design for maximum deflection sensitivity would be one in which the deflecting electrodes are curved and spaced so that the edge of the beam just grazes the deflecting electrodes at the maximum deflection angle. In practice, this condition may be closely approximated by mounting the electrodes at an angle with respect to the gun axis and bending the electrodes at one or two places.

The capacitances between deflecting electrodes and between the deflecting electrodes and other gun electrodes are important considerations in cathode-ray oscilloscopes employing wide-band deflection amplifiers. High-frequency response of the deflection amplifier can be achieved more readily if the oscillograph tube has high deflection sensitivity and low interelectrode capacitances.

The capacitance between the deflecting electrodes of a pair can be minimized by reducing the width of the deflecting electrodes and by mounting the electrodes so that the supports do not contribute appreciable capacitive coupling. Another important consideration is the capacitance between stem leads. If possible, stem leads adjacent to deflecting-electrode leads should be left open.

The deflecting electrodes closer to the cathode (DJ3 and DJ4, Fig. 2) require less width than the top electrodes (DJ1 and DJ2, Fig. 2) because the beam is deflected in the plane perpendicular to the lower deflecting electrodes which passes through the gun axis. The minimum width of these electrodes should be 1.5 times the distance between the electrodes. If the electrodes are narrower, edge effects will cause a nonuniform deflecting field and a loss of deflection sensitivity may result.

DEFLECTING-ELECTRODE SHIELD

An electrostatic shield is placed between the two pairs of deflecting electrodes. This shield usually consists of a slotted plate mounted so that the center of the slot is on the tube axis and the long dimension is in the same direction as the deflection produced by the DJ3-DJ4 electrodes (Fig. 2). The deflecting-electrode shield

reduces the capacitive coupling between the two pairs of deflecting electrodes. The shield is usually connected within the tube to grid No. 4. When high-frequency deflecting voltages and high-impedance deflecting-electrode coupling circuits are used, the capacitive coupling has the greatest effect, and effective shielding is needed.

DEFLECTING-ELECTRODE DESIGN

Center of Deflection

The center of deflection of a pair of deflecting electrodes is the point on the gun axis that intersects the backward projection of the path of a deflected electron beam after the beam leaves the influence of the deflecting field (Fig. 3). For a flat, parallel pair of deflecting electrodes, the center of deflection is on the axis at a distance from the edge closest to the cathode of the pair of deflecting electrodes equal to 1/2 the length of the electrodes.

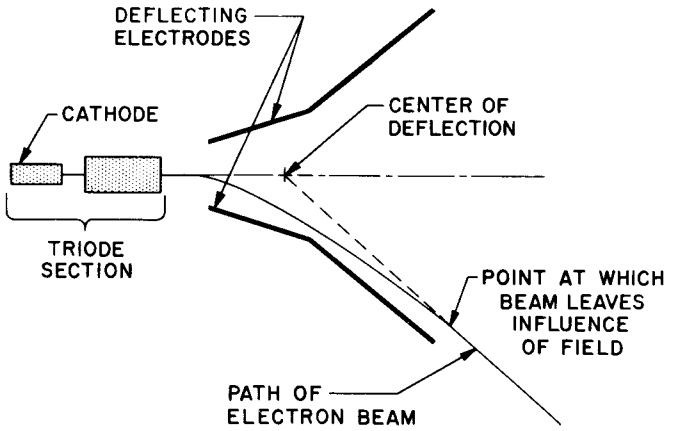


Figure 3. Center of Deflection of a Pair of Deflecting Electrodes

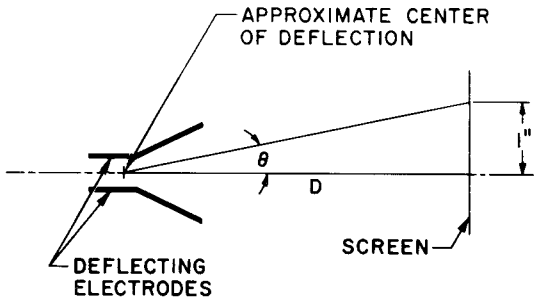
For bent deflecting electrodes and those spaced at an angle with respect to the tube axis, the center of deflection is not as easily determined. For most nonparallel electrodes, the center of deflection may be assumed to be approximately 1/3 the length of the electrodes from the edge of the electrode closest to the cathode. This point may be more accurately determined by geometrically constructing the path of the electron beam (as discussed in the following text) and projecting backward to find the center of deflection.

Deflection Sensitivity

Deflection sensitivity is defined as the number of degrees of electron-beam deflection that will occur per volt of potential difference across the deflecting electrodes. Deflection sensitivity DS can be calculated from the relation (see Fig. 4):

$$DS = \frac{\cot^{-1} D \text{ (degrees)}}{\text{deflection factor (volts/inch/kilovolt)}}$$

where D = the distance from the center of deflection to the screen (inches)



$$\cot \theta = D/l$$

$$\theta = \cot^{-1} D$$

$$DS = \frac{\cot^{-1} D}{\text{DEFLECTION FACTOR (VOLTS/INCH/KILOVOLT)}}$$

Figure 4. Determination of Deflection Sensitivity

The deflection sensitivity of a proposed design can be accurately determined by comparing the proposed design to a similar design for which data are available. To permit the comparison to be made, divide the deflecting electrodes of both the known design and the proposed design into several lengths which will be treated as incremental, parallel-plate electrodes. The average distance between each incremental pair (midpoint distance) is considered to be the distance between the equivalent parallel-plate electrodes (see Fig. 5).

The deflection sensitivity for the proposed design can be found by means of the relation:

$$DS' = DS \frac{l'_1/d'_1 + l'_2/d'_2 + l'_3/d'_3 + \cdots}{l_1/d_1 + l_2/d_2 + l_3/d_3 + \cdots}$$

where DS' = deflection sensitivity of the proposed design

DS = deflection sensitivity of the known design

$l'_1, l'_2, l'_3, \ldots$ = lengths of the incremental parallel-plate deflecting electrodes of the proposed design

$d'_1, d'_2, d'_3, \ldots$ = distances between the incremental parallel-plate electrodes corresponding to $l'_1, l'_2, l'_3, \ldots$ for the proposed design

$l_1, l_2, l_3, \ldots$ and $d_1, d_2, d_3, \ldots$ = lengths and distances corresponding to $l'_1, l'_2, l'_3, \ldots$ and $d'_1, d'_2, d'_3, \ldots$ for the known design

The accuracy of the value of deflection sensitivity obtained by the preceding method increases as the number of increments used is increased. Six to eight increments are usually sufficient.

Electron-Beam Trajectory

After the deflection sensitivity of the proposed design has been found, it is necessary to plot the path of the

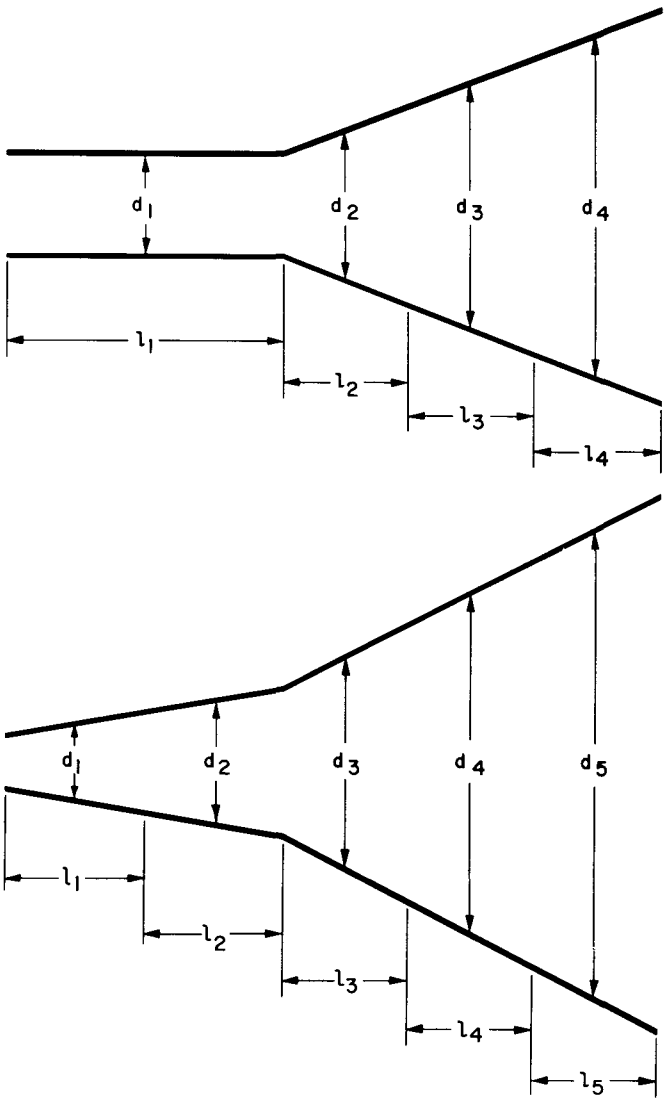


Figure 5. Two Examples of the Method of Dividing Nonparallel Deflecting Electrodes into Incremental Parallel-Plate Electrodes

electron beam at maximum deflection to be sure that the beam will not strike the deflecting electrodes. The procedure is as follows:

1. Determine the angle between the center line of the gun and a line from the approximate center of deflection to the point of maximum beam displacement on the screen. (This angle is one-half the maximum deflection angle.)

2. Use the incremental parallel-plate electrode approximation that was constructed to determine deflection sensitivity and find the contribution of each increment to the half-maximum-deflection angle by means of the relation:

$$\text{incremental deflection angle} = \text{1/2 maximum deflection angle} \cdot \frac{l_1'/d_1'}{l_1'/d_1' + l_2'/d_2' + l_3'/d_3' + \cdots}$$

3. Plot an enlarged diagram of the proposed deflecting-electrode design with the spacings and length drawn to scale as shown in Fig. 6. Show also the midpoints of the increments chosen above.

4. Draw a line, at angle θ_1 (Fig. 6), with respect to the gun axis, which passes through the point on the axis which is the midpoint of the first increment. At the point this line intersects the line perpendicular to the axis through the midpoint of the second increment, draw a second line at an angle θ_2 with respect to the first line. Continue this procedure for all increments.

5. The line plotted indicates the approximate path of the center of the electron beam. To determine whether the beam will strike the plate, consider that the beam has a diameter determined by the masking aperture and plot the beam boundaries with relation to the center of the beam.

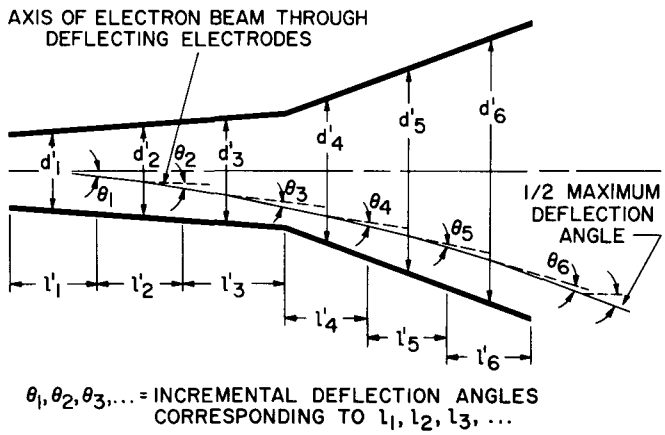


Figure 6. Electron Beam Trajectory at Maximum Deflection

POST-DEFLECTION ACCELERATION

Post-deflection acceleration is used in some types of electrostatic-deflection tubes to increase the brightness of the display by applying to the post-accelerator electrode a voltage higher than that of the ultor. Application of the higher voltage is accomplished by separating the bulb wall coating and applying the post-accelerating voltage to the coating section (post-ultor) nearest the screen. Fig. 7 illustrates a tube that uses post-deflection acceleration.

Because the deflecting electrodes operate on the beam before post-acceleration, deflection sensitivity is little changed by the higher voltage applied to the post-ultor. The post-ultor voltage does, however, result in a focusing action which bends the deflected beam toward the tube axis. As a result, in a typical tube type utilizing a post-ultor voltage of twice the ultor voltage, the deflection voltage must be increased by about 20 per cent to obtain the same deflection angle as would be obtained with the same ultor voltage alone. For single-band post-accelerator tubes, a voltage step-up ratio of 2:1 is usually used. Higher ratios result in excessive pattern distortion. Satisfactory operation can be obtained with a ratio as high as 10:1 by using multiple-band post-accelerator coatings or helical

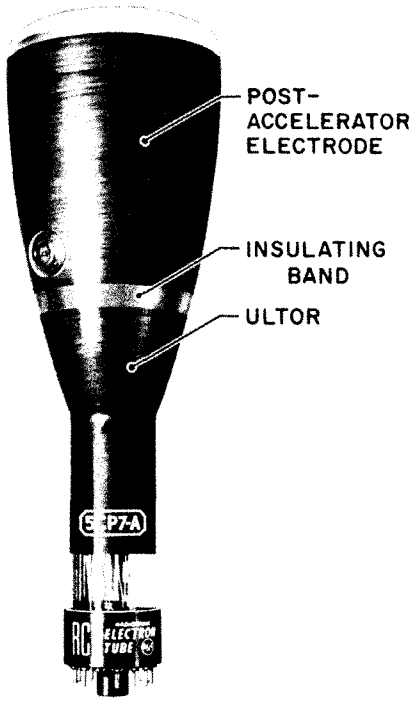


Figure 7. Oscillograph Tube That Uses Post-Deflection Acceleration

coatings.

Post-deflection acceleration offers the advantage, in addition to increased brightness and lower deflecting-voltage requirements, that the high-voltage power supply may be grounded between the negative and positive terminals, thereby permitting the use of filter capacitors having lower voltage ratings, and reducing the problems associated with high-voltage breakdown. In many applications, these advantages justify the use of the more expensive post-deflection-accelerator types.

The use of post-deflection acceleration causes the edge focus to show some deterioration as compared to the performance without post-deflection acceleration. The explanation is that the beam in a tube having post-deflection acceleration must be deflected through a wider angle at the deflecting electrodes to obtain the same screen deflection as would occur in the tube without post-deflection acceleration. There is also a loss in deflection sensitivity due to the fact that the post-accelerator type must employ wider spaced deflecting electrodes to make the wider deflection angle feasible.

In post-accelerator types, the secondary electrons emitted from the gun structure (for example, secondary electrons resulting from bombardment of the deflecting electrodes by the beam at wide deflection angles) are accelerated to the phosphor screen by the post-ultor voltage and cause the screen to glow in a diffuse halo, thereby reducing the contrast of the display.

MAGNETIC DEFLECTION

Basic design considerations for magnetic-deflection oscillograph tubes are similar to those outlined for electrostatic-deflection types.

The general design of the magnetic-deflection guns is described in detail in the chapter on picture tube design. In some cases, special features are necessary to fit particular applications. For example, projection tubes and flying-spot tubes are usually operated at considerably higher ultor voltages than other types and thus require special designs to prevent voltage breakdown. Fig. 1C shows a typical gun designed for this type of operation. Note that grid No. 3 is the focusing electrode. To prevent voltage breakdown between grid No. 3 and grid No. 4 (the bulb coating), it is necessary to have a smooth, clean surface on the grid No. 3 thimble. The thimbles are, therefore, cleaned and polished to eliminate sharp points which might cause an increase in the voltage gradient sufficient to cause grid emission.

In flying-spot tubes and tubes for recording TV pictures, the beam current requirements are usually considerably less stringent than those for picture tubes. To improve resolution, these types utilize smaller masking apertures and longer guns than are used for other oscillograph types.

PHOSPHOR SCREEN

Different phosphor screens are used in oscillograph tubes to meet the requirement of various applications. Usually, a tube type is available in several different phosphor screens. For example, the 5U type is produced with the P1 (5UP1), P7 (5UP7), and P11 (5UP11) phosphors.

Information concerning color, visual efficiency, persistence, spectral emission, and efficiency of energy conversion is available for most registered phosphors. These characteristics should be considered in determining the phosphor best suited for a particular application.

Aluminum backing of phosphors is used to increase the brightness and improve the contrast of oscillograph tubes. Aluminizing is advantageous only if the screen voltage exceeds 5 kilovolts, otherwise the energy lost by the beam electrons in passing through the aluminum more than offsets the gain in brightness produced by light reflected from the aluminum backing.

In projection tubes and other types operating at ultor voltages greater than 18 kilovolts, an aluminized screen has an additional purpose. Phosphors have a second secondary-emission cross-over potential lower than the ultor voltage applied to projection types. If the screen were not aluminized, the phosphor potential would not rise above the secondary emission cross-over potential and therefore no increase in brightness would occur for an increase in applied voltage from 18 kilovolts to 40 kilovolts.

PROCESSING

The processing schedules used for oscillograph-tube

production are the result of years of development and production experience. They are determined from considerations of the quality of the resulting product, the cost of the process, and the cost of the scrap that may be expected should the process be omitted or another substituted having different cost.

Changes in some processing schedules occur frequently. For example, grid emission from grid No. 1 aperture may cause scrap in some tube types. When grid emission occurs, the processing of the aperture may be changed to include gold plating, which is one method of reducing grid emission. Because gold plating increases the cost of the part, tests may be made periodically to determine if the gold plating is still necessary. Frequently, the later tests show that grid emission is no longer a problem and that unplated apertures can be used successfully. However, the majority of processing schedules are relatively permanent. Changes are initiated primarily to improve quality or to lessen cost without sacrifice of quality.

The following enumeration indicates some of the processing schedules used in the production of an oscillograph tube. A washing process, may actually be any one of literally hundreds of different washing schedules now in use. The type of wash specified is determined by the material, the finished tube requirements, the facilities available at a given location, comparative costs, and many other factors. Similar comments apply to the other processes listed. Numerous inspections of parts and assemblies are specified in the complete processing specifications, but have not been included in the list.

Bulb

1. Wash
2. Apply screen
3. Coat exterior of bulb
4. Bake in air

Gun Parts (Stainless Steel)

1. Clean, de-burr, polish, etc.
2. Wash
3. Fire in dry hydrogen

Gun Assembly

1. Clean stems
2. Air bake (ceramic mounts)
3. Seal-in bulb

Exhaust (Rotary exhaust machine)

1. Bake
2. Pre-flash getters
3. Activate cathode
4. Flash getters
5. Tip off tubulation

Aging

1. Hot-shot treatment
2. Age

BIBLIOGRAPHY

Moss, Hilary, "Engineering Methods in the Design of the Cathode Ray Tube," Jour. Brit. IRE, Vol. 5, No. 5, Oct.-Dec., 1945

Moss, Hilary, "The Electron Gun of the Cathode Ray Tube (Parts 1 and 2)," Vols. 5 and 6, Nos. 1 and 2, Jour. Brit. IRE, Jan.-Feb., 1945 and March-May 1946

Maloff, I. and D. W. Epstein, Electron Optics in Television, McGraw-Hill Book Co., 1938

Sponsler, G. C., "An Automatic Electron-Trajectory Tracer Study of Helix and Band-Type Post Deflection Acceleration," Lincoln Lab Tech. Rpt. 64

White, W. G., "Cathode Ray Tubes with Post Deflection Acceleration," Electronic Engineering, No. 21, p. 75, 1949

Leverenz, H. W., An Introduction to Luminescence of Solids, John Wiley and Sons, 1950

Leverenz, H. W., "Final Report on Research and Development Leading to New and Improved Radar Indicators," PB25481 and PB32546, Office of the Publication Board, Department of Commerce, Washington 25, D.C.

"RCA Phosphors for Cathode-Ray Excitation," RCA Electron Tube Div. Harrison, N. J., Catalog No. TPM-1508, 1957

Phototube Design

J. L. Weaver

Lancaster

This brief review of phototube technology emphasizes the design of multiplier phototubes, because the design of single-unit or of twin-unit vacuum and gas phototubes involves similar techniques. Typical photocathodes are considered first, with emphasis on the requirements of multiplier phototubes. Representative types of secondary-emission multipliers are described. The overall performance of multiplier phototubes is discussed in relation to present design tests.

PHOTOCATHODES

Photocathodes of commercial interest are electrodes capable of emitting electrons into a vacuum when exposed to light or similar electromagnetic radiation. The rate at which electrons are emitted from a photocathode is proportional to the intensity of the incident radiation over a range of at least ten billion to one. The maximum velocity v with which the electrons are emitted is related to the frequency of the radiation by the Einstein equation,

$$\frac{mv^2}{2} = h\nu - \phi$$

where m is the mass of an electron
h is Planck's constant
ν is the frequency of the radiation
ϕ is the photoelectric work function of the photocathode

Photocathodes which are sensitive to visible radiation invariably contain one or more of the alkali metals. In comparison to most other elements, these metals have the lowest photoelectric work functions. The lowest known value is 1.8 electron-volts (ev) for cesium. The pure alkali metals are inefficient photoelectric emitters because of their high optical reflectivity and strong electron affinity. Semiconductors containing these elements, however, not only have high photoelectric yields but also often have lower photoelectric work functions than the pure metals.

Present photocathodes of the semiconducting type have, as an essential constituent, either the oxide of an alkali metal or an intermetallic compound of an alkali metal and a Group V element such as antimony. The silver-oxygen-cesium photocathode, for example, consists chiefly of amorphous cesium oxide containing colloidally dispersed silver and a stoichiometric excess of cesium.[1] This cathode is sensitive to radiation of wavelengths between 0.3 and 1.2 microns, but if a silver-bismuth alloy is used in place of the silver the response to visible radiation is greatly increased and infrared response is negligible.[2] Rubidium may be substituted for cesium, as in the silver-oxygen-rubidium photocathode.

The cesium-antimonide photocathode[3] is an example of the intermetallic-compound type of cathode. In this cathode cesium antimonide, Cs_3Sb, is the compound associated with photoelectric emission. Maximum emission is obtained with blue and ultraviolet radiation; practically none, with infrared radiation. Other photocathodes of this type are the multialkali cathodes, $(NaK)_3Sb$ and $(CsNaK)_3Sb$. Characteristics of these new photocathodes are described in the literature.[4, 5, 6] Although Cs_3Bi, another photoemissive intermetallic compound, has a much lower sensitivity than that of Cs_3Sb, it has been used in multiplier phototubes intended for use in colorimeters.

Photocathodes are formed either on a metal electrode, as a thin semitransparent layer on the glass wall or faceplate of the tube, or as a semitransparent layer on a reflecting metallic substrate. Contact to the semitransparent layer on the glass wall or faceplate is usually obtained by overlapping the semitransparent layer on an opaque aluminum film to which the cathode lead is connected. Conductivity of semitransparent cathodes deposited on glass is important because the cathode current must flow laterally through the thin layer. Conductivity of cesium-antimonide, for example, decreases exponentially by a factor of ten for each 55 C decrease in temperature below 40 C.[7]

Conductivity across the semitransparent cesium-antimonide photocathode can be increased by several methods. One method is to deposit that photocathode on an open grill of opaque aluminum which has been previously deposited on the glass. The grill provides bus bars for the photocurrent and thereby lowers the voltage drop in the cathode material for a given photoemission current. Another method is to deposit the photocathode on semitransparent backing layers, or substrates, of inactive metals such as gold or rhodium. A disadvantage of this method is that such a layer absorbs 10 to 30 per cent of the incident radiation. This can be overcome by using a transparent layer of stannic oxide as a substrate. The conductivity of this material,

however, decreases as the temperature is lowered, thereby reducing the effectiveness of this method.

Low-temperature operation of multiplier phototubes is desired in many applications because it reduces the background noise level due to thermionic emission. At 20 C, thermionic emission from cesium-antimonide is of the order of 10^{-15} ampere/cm^2 as compared to 10^{-9} to 10^{-12} ampere/cm^2 for the Ag-O-Cs cathode. The thermionic emission decreases almost exponentially with temperature over a wide range. Because photoelectric emission is practically unimpaired by cooling the cathode to low temperatures, an appreciable improvement in signal-to-noise ratio is thus obtained.

The sensitivity of a photocathode may be expressed in several different ways. For practical measurements cathode "luminous sensitivity," expressed in microamperes of photocurrent per lumen of incident luminous flux, is a useful figure of merit. Luminous flux is customarily obtained from a tungsten lamp operating at a color temperature of 2870 K. Because most of the radiation from this illuminant lies in the red and infrared regions of the spectrum, this method tends to overrate the sensitivity of red-sensitive phototubes in comparison to that of blue-sensitive types.

A complete description of cathode sensitivity is given by the spectral sensitivity characteristic, which shows the photocurrent emitted per unit of incident, monochromatic, radiant power as a function of wavelength. The sensitivity at any particular wavelength is called "radiant sensitivity" and is frequently expressed in amperes per watt. This unit is analogous to the number of photoelectrons emitted per electron-volt of incident energy.

The energy q per photon of radiation of wavelength is given by the quantum relation $q=hc/\lambda$, where h = Planck's constant and c = velocity of light. If λ is given in angstroms and q is expressed in electron-volts, then hc = 12,400 ev-angstroms. Thus, the ratio of the number of electrons emitted to the number of photons incident, or the "quantum efficiency" of a photocathode at a given wavelength λ, is the product of radiant sensitivity at this wavelength expressed in amperes per watt and the factor $12{,}400/\lambda$. The peak quantum efficiency of cesium-antimonide cathode is about 12 per cent at 4000 angstroms. The highest peak quantum efficiencies of commercial phototubes are 20 to 25 per cent.

Spectral sensitivity characteristics of typical photocathodes in phototubes are shown on semilog plot in Fig. 1. Average characteristics of corresponding cathodes are listed in Table I. Because radiant sensitivity is directly proportional to luminous sensitivity for a given spectral sensitivity characteristic, the ratio of the two may be used to compute the radiant sensitivity of a cathode from a measurement of its luminous sensitivity.[8]

Several aids to the study of known photocathodes and to the "design" of new ones are available. The energy-band structure of a cathode may be determined from measurements of its spectral sensitivity characteristic, its optical absorption characteristic, the behavior of its conductivity and photoconductivity as functions of temperature and, possibly, from Hall-effect data. Chemical analysis of actual photocathodes is difficult because of the small quantities present. Trace impurities and small changes in chemical composition greatly influence the photoelectric characteristics. However, mass-spectrometer and radioisotope techniques may eventually yield significant information, particularly in regard to cathode contamination problems.

SECONDARY-ELECTRON MULTIPLIERS

Because the photoelectric current produced at low light levels is very small, current amplification within the phototube is highly desirable in commercial phototubes. Amplification of photoelectric current by factors of 4 to 10 is readily obtained by emission-limited ionization of the inert gas contained in a gas phototube. However, gas amplification of the photocurrent limits the speed of response to that of ion transit time, which is of the order of 10^{-4} second in gas phototubes containing argon. Amplification of the photocurrent by means of secondary emission, on the other hand, provides fast response to modulated light as well as very high current-amplification factors.

In a multiplier phototube several secondary-emitting electrodes, or dynodes, are used in cascade. Secondary electrons from one dynode are accelerated and directed toward the succeeding dynode. Successive dynodes are connected to external potentials which increase by 100 volts or more per dynode stage.

Multiplier Dynodes

Typical surfaces which have a high secondary-emission ratio δ at reasonably low voltages are cesium antimonide (Cs_3Sb), oxidized silver-magnesium (AgMgO), oxidized copper-beryllium (CuBeO), and silver-oxygen-cesium ($AgCs_2O$). The cesium antimonide and silver-oxygen-cesium surfaces are similar to the corresponding photocathodes previously described. The AgMgO surface[9] consists of a thin layer of MgO formed on silver alloy containing a few per cent of magnesium. The CuBeO surface is similar.[10] The MgO and BeO are insulators from which internally excited secondary electrons can escape. However, these insulating layers must be thin enough or have sufficient impurity conductivity to permit electrons from the underlying metal to restore charge neutrality. Cesium antimonide is a semiconductor which has a high density of impurity centers from which secondary electrons can be excited. The high secondary-emission ratio of $AgCs_2O$ is attributed to the excess cesium which the surface invariably contains. Cesium absorbed on a secondary-emitting surface tends to reduce its electron affinity and thereby facilitates escape of internal secondary electrons.[11]

The average current amplification per stage μ_o is shown in Fig. 2 as a function of voltage per stage. In a well-designed multiplier μ_o is approximately equal to the secondary-emission ratio δ. This graph will be used later in the text in the discussion of multiplier phototube testing. Characteristics of the surfaces are

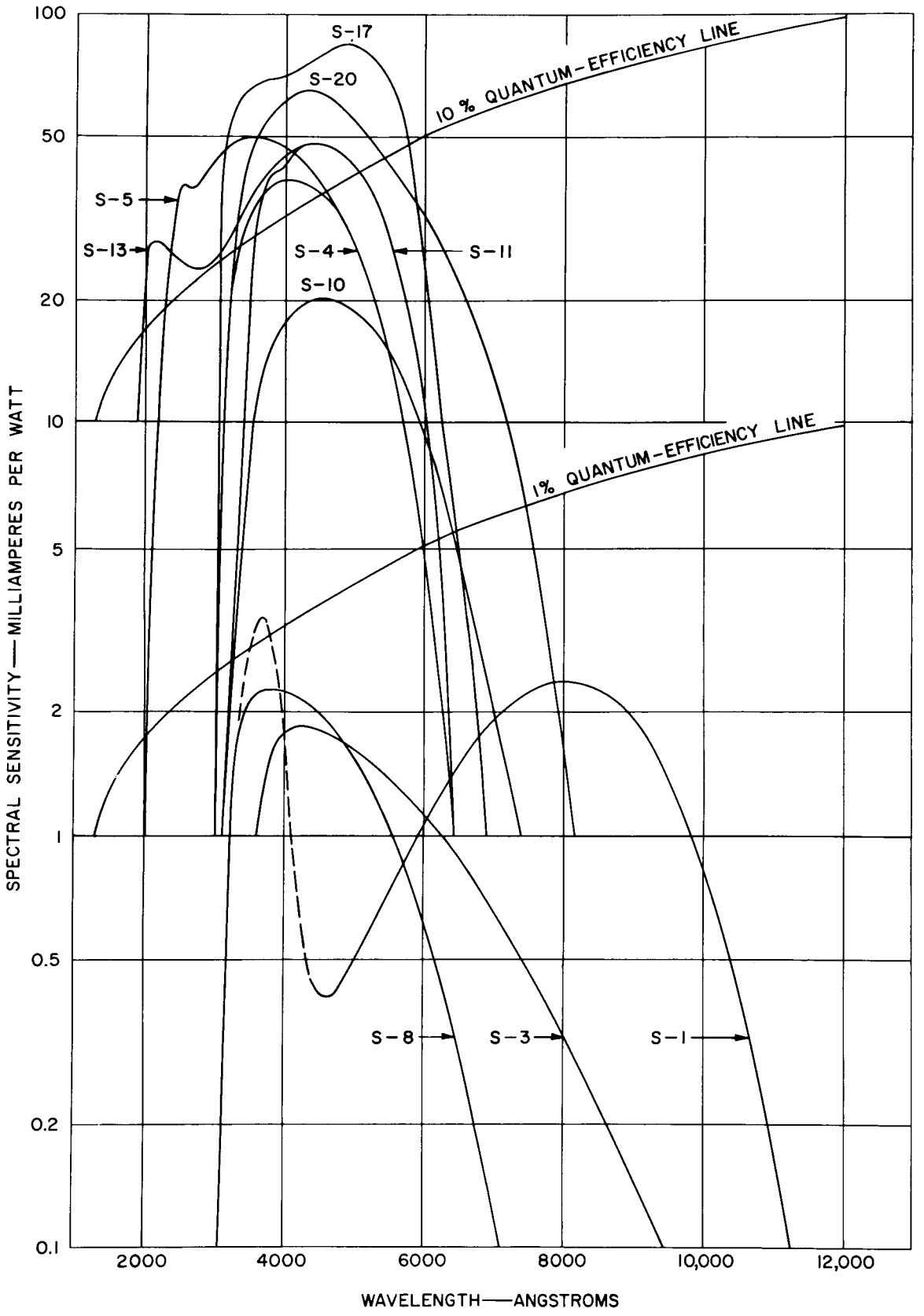


Figure 1. Spectral Sensitivity Characteristics of Photocathodes in Phototubes

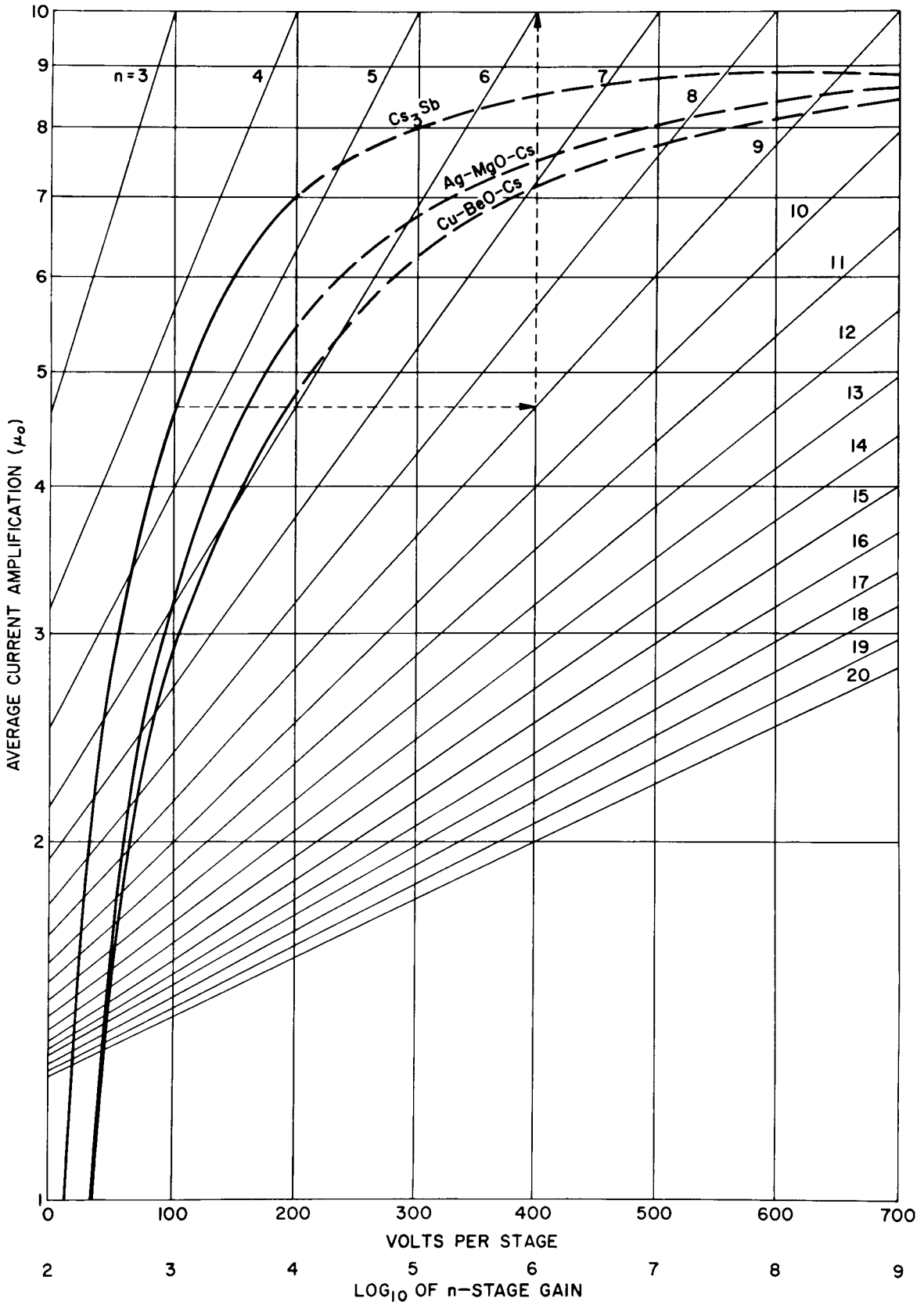


Figure 2. Average Current Amplification per Stage as a Function of Voltage per Stage

Table I

Typical Cathode Characteristics*

Device S-Number**	Photocathode Type and Envelope	Wavelength of Maximum Response (angstroms)	Radiant Sensitivity*** (ma/watt)	Quantum Efficiency*** (per cent)	Luminous Sensitivity**** (μa/lumen)
S-1	Ag-O-Cs Lime-glass bulb	8000	2.35	.36	25
S-3	Ag-O-Rb Lime-glass bulb	4200	1.86	.55	6.5
S-4	Cs-Sb Lime-glass bulb	4000	39.1	12	40
S-5	Cs-Sb 9741 glass bulb	3400	50.1	18	40
S-8	Cs-Bi Lime-glass bulb	3650	2.26	.77	3
S-10	Ag-Bi-O-Cs Semi-Transparent, Lime-glass bulb	4500	20.3	5.6	40
S-11	Cs-Sb Semi-Transparent, Lime-glass bulb	4400	48.2	14	60
S-13	Cs-Sb Semi-Transparent Fused-silica bulb	4400	47.7	13	60
S-17	Cs-Sb Lime-glass bulb, Reflecting Substrate	4900	83.0	21	125
S-20	Sb-K-Na-Cs Semi-Transparent, Lime-glass bulb	4200	64.2	18	150

* These are typical values obtained in practice. For maximum observed values see the article "Photoelectric and Secondary Electron Emission" by A. H. Sommer in this book.

** These characteristics are shown in Fig. 1. They refer to spectral responses of typical phototubes rather than to photocathodes, and include the effect of the tube envelope.

*** Values given are for wavelength of maximum response.

**** Light source is a tungsten-filament lamp operated at a color temperature of 2870 K.

dependent upon the forming process. For example, the characteristics shown are those which are obtained commercially when the surfaces are exposed to cesium vapor. Without cesium activation the values of secondary-emission ratio δ are somewhat lower.

To be usable in multiplier phototubes a secondary-electron emitter should: (1) have a high secondary-emission ratio, (2) be compatible with the photocathode, and (3) show negligible change in emission during tube life. Oxidized silver-magnesium and copper-beryllium show less fatigue under electron bombardment than does Cs_3Sb, but their values of δ are lower than that of Cs_3Sb. The high thermionic emission associated with silver-oxygen-cesium dynodes prohibits their use in high-gain multipliers. The development of new or improved secondary-electron emitters for multiplier phototubes involves techniques similar to those employed in the development of a new photocathode.

Multiplier Structures

Phototubes containing unfocused multiplier structures are shown in Figs. 3A and 3B. Fig. 3A shows the construction of the "Venetian blind" type of multiplier, while Fig. 3B shows a multiplier that uses a box type of dynode. Accelerating grids are used to direct electrons from dynode to dynode. In both structures, electron transit time between stages varies over a relatively wide range as a result of the range in length of trajectories.

Electrostatically focused multipliers[12] are shown in Figs. 4A and 4B. In these structures the electric field between consecutive dynodes is given a configuration which focuses the space current between stages. With properly designed dynodes, electrons emitted from a relatively large area on one dynode can be focused upon a smaller area on the succeeding dynode. The dynodes are usually mounted between flat insulating spacers which support the dynodes and confine the space current.

If an electrostatically focused multiplier is used in conjunction with a large photocathode, the photoelectrons must be accurately focused upon the effective area of the first dynode. The effective area is that portion of the dynode surface from which secondary electrons are focused upon the succeeding dynode. In the structures of Figs. 4A and 4B the effective area

may be a little more than half of the active dynode surface because part of the dynode is used to shape the interdynode field. Satisfactory focusing of photoelectrons on the first dynode is accomplished in these structures by means of an auxiliary focusing electrode.

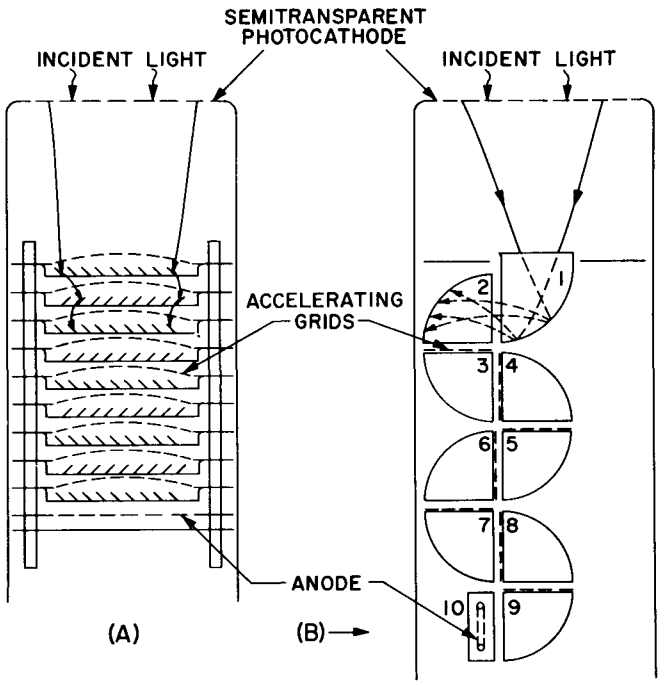


Figure 3. Phototubes Containing Unfocused Multiplier Structures: (A) "Venetian blind" type of multiplier; (B) box type multiplier

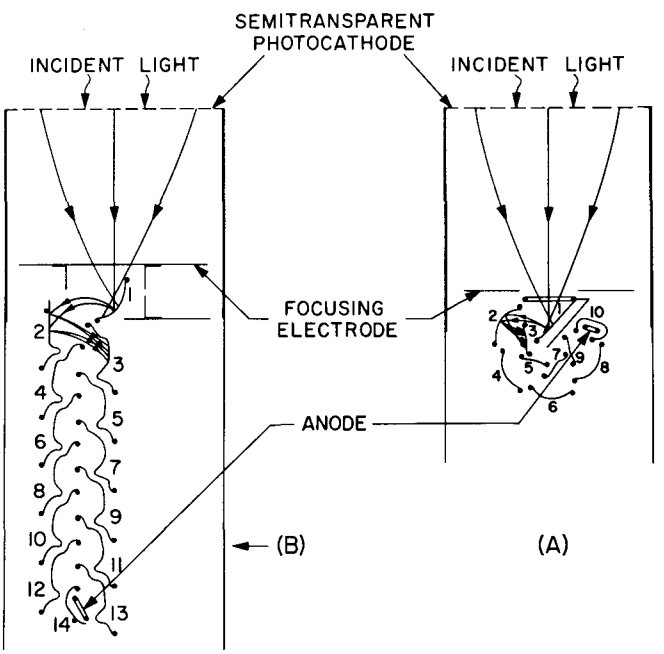


Figure 4. Electrostatically Focused Multipliers: (A) circular multiplier structure used in RCA 6655; (B) in-line multiplier structure used in RCA 6810A

An electromagnetically focused multiplier is shown in Fig. 5. A transverse magnetic field (perpendicular to the page) causes the electrons to follow cycloidal trajectories between dynodes.[13] This type of multiplier requires a uniform magnetic field of large cross section. It should be noted that the structure shown here permits feedback of ions from anode to cathode. Ion feedback is minimized in electrostatically focused multipliers by overlapping consecutive dynodes to obstruct possible ion paths between anode and cathode.

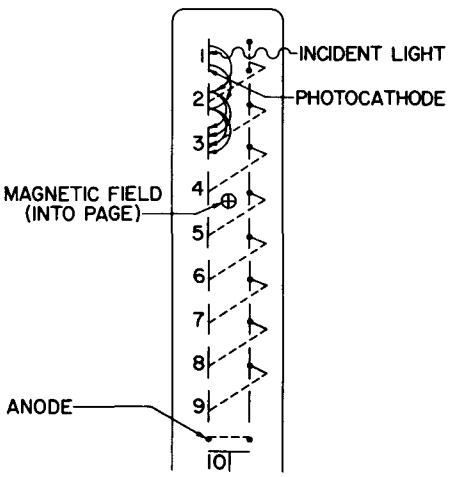


Figure 5. Electromagnetically Focused Multiplier Structure

A transmission type of multiplier is shown in Fig. 6.[14] The secondary-emitting material is deposited on a very thin metal film which scatters the high-velocity primary electrons. The scattered primaries cause secondary electrons to be emitted from the active material on the opposite side of the metal film. A strong plane-parallel field between consecutive dynodes minimizes electron scattering due to initial velocities of the secondaries and assures small spread in their transit time between stages. In these respects the transmission type of multiplier is almost ideal. However, high interstage voltage (approximately 2 kilovolts minimum) is required to obtain values of $\mu_0 \gtrsim 2$. Development of ultra-thin metal films and stable, thin-film secondary emitters supported on a fine metal mesh may eventually make the transmission type of multiplier commercially feasible.

The current output of a multiplier is collected by the anode, which is connected through the external load impedance to a potential about 100 volts more positive than that of the last dynode. The anode design should be such that: (1) fluctuations of anode potential do not defocus the latter stages of the multiplier, and (2) all of the current from the last dynode is collected by the anode over a wide range of voltage between the last dynode and anode. These requirements are met in the multipliers shown in Figs. 3 to 6 by using an open grid structure as anode. Current from the second to the last dynode passes through the open grid structure to the last dynode; secondary electrons from this dynode are then collected on the anode. By means of this close-spaced geometry large space-charge-limited currents are possible under pulse operation.

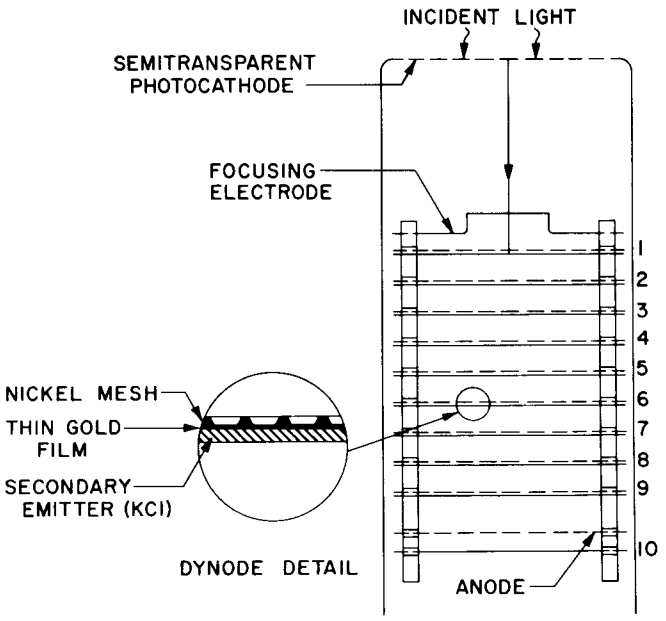


Figure 6. Transmission Type of Multiplier

Several techniques are available for the design of multiplier structures. A rapid and fairly reliable method for investigating the performance of electrostatic types of multipliers is the rubber-membrane analog.[15, 19] If the multiplier structure has a uniform cross section along one axis, the electric field between dynodes is simulated by the slopes of a rubber membrane stretched over a scale model of the multiplier cross section. The elevation of each model dynode is made proportional to its negative potential. For small slopes of a uniformly elastic membrane it can be shown that contours of equal elevation correspond to equipotentials in the multiplier. Furthermore, a small ball rolling down the slopes between consecutive dynodes on the model follows a path which approximates the trajectory of electrons in the multiplier. With appropriate modifications, transit time, ion trajectories, and the effects of initial velocity of the secondary electrons can also be estimated from the model. Special techniques may be used for simulation of space charge,[16] and for the analog corresponding to axial symmetry.[17]

A more accurate analysis of multiplier performance can be obtained from an accurate plot of equipotentials which, in electrostatically focused multipliers, are solutions of Laplace's equation. This equation is difficult to solve for practical structures because of the intricate boundary conditions. The equipotentials can be determined mathematically by relaxation methods (successive approximation),[18] and experimentally by means of the electrolytic-tank[19] or resistance-network analog.[20] After the equipotentials are known electron trajectories can be plotted graphically.[19, 21]

Actual electron trajectories can be estimated by coating the dynodes with a phosphor. When the multiplier is operated at one to two kilovolts per stage, sufficient current amplification is obtained to display the areas of electron incidence through several stages of the multiplier. Focusing of photoelectrons on the first dynode can be observed in the same manner. Another technique for multiplier design is to use dynodes which are mounted adjustably in a demountable exhaust system. A thermionic cathode is used as an electron source.

PHOTOTUBE DESIGN TESTS

Vacuum and Gas Photodiode Tests

Fundamental tests by which vacuum and gas phototube designs are evaluated are: (1) cathode sensitivity, both luminous and radiant as previously described; (2) dark current; and (3) anode current as a function of anode voltage and light flux input. Dark current is the current which flows from the output electrode when all radiation to which the tube is normally sensitive is excluded. In addition to thermionic emission from the photocathode the dark also includes electrical leakage current across the insulation between anode and cathode terminals. Leakage current is held to a minimum in well-designed phototubes by providing a long leakage path between terminals and by avoiding deposition of conducting material between the terminals.

Anode current of a typical vacuum phototube is shown as a function of anode voltage in Fig. 7. The slope of the curves above 50 volts is indicative of the residual gas pressure in the tube.

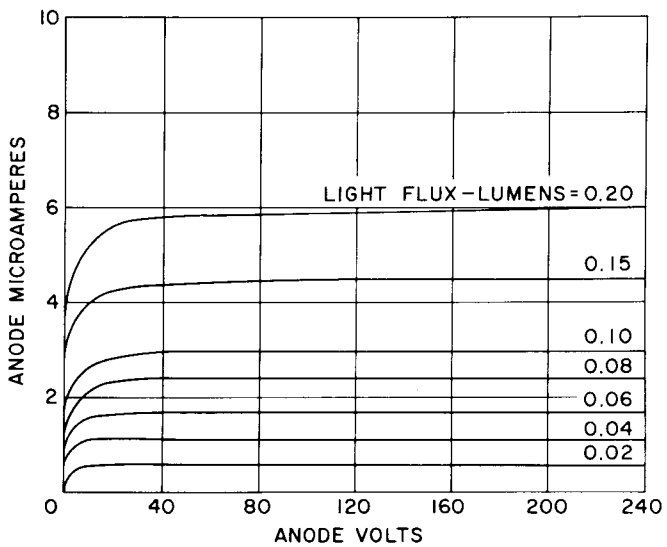


Figure 7. Anode Characteristics of a Vacuum Phototube (type 934)

Anode characteristics of a gas phototube are shown in Fig. 8. Above 15 to 25 volts the anode current increases in proportion to the gas amplification factor of the tube. For test purposes gas amplification factor is defined as the ratio of sensitivity at 90 volts to that at 25 volts. At a voltage at least 5 per cent greater than the maximum rated supply voltage, the anode current increases abruptly to excessive values as a result of a self-sustained glow discharge.

Multiplier Phototube Tests

Evaluation of multiplier phototubes requires similar

fundamental tests. Sensitivity of the photocathode is measured by connecting all dynodes together and testing the tube as a photodiode with all the dynodes serving as the anode of the photodiode. Anode sensitivity of the multiplier phototube is measured at a reduced light level and with dynode supply voltages fixed at some specified value. The ratio of anode sensitivity to cathode sensitivity is the "current amplification factor" of the multiplier. Current amplification of the individual dynode stages is measured occasionally during the design of a multiplier as a check on its secondary emission and electron-focusing characteristics. The current amplification of a dynode is very nearly equal to its secondary-emission ratio if electrons are properly focused from dynode to dynode. Over the normal range of interstage voltage, the average current amplification can be approximated as a function of voltage by the relation:

$$\mu_o = cV^p \qquad (1)$$

where c = a constant,
V = supply voltage per stage
p = a constant ≈ 1 (≈ 0.7 for Cs_3Sb and ≈ 0.9 for AgMgO)

Thus, for n similar dynodes in cascade, the overall current amplification $\mu_n = c^n V^{pn}$. Values of log μ_n for n stages of identical dynodes, each having a current gain μ_o are given in Fig. 2. For example, a nine-stage multiplier using cesium-antimonide dynodes and operating at 100 volts per stage will have a current amplification of 4.65 per stage and an overall gain of 10^6. Because the overall current amplification varies as (total supply voltage)^{pn}, regulated high-voltage supplies are recommended for tests and commercial applications of multiplier phototubes.

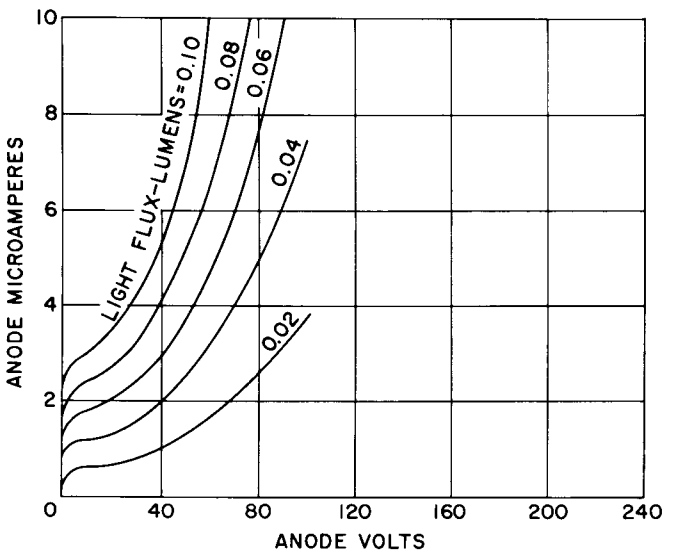


Figure 8. Anode Characteristics of a Gas Phototube (type 5583)

Several sources contribute to the dark current which is measured at the anode of a multiplier phototube.[22] Ohmic leakage between base pins and across the dynode spacers and glass stem accounts for dark current which is proportional to the overall supply voltage. An excess of free cesium condensed on the insulation in the tube can cause low internal leakage resistance in vacuum and gas phototubes as well as in multiplier types.

At normal ambient temperatures, thermionic emission from the photocathode is the major cause of dark current in properly designed multiplier phototubes. Since this emission is a characteristic of the cathode, it can be reduced only by cooling the tube to a low temperature, by making the cathode area small, or by permitting electrons from only a small portion of the total cathode area to enter the multiplier.

Two other possible causes of dark current in multiplier phototubes are ion feedback and photon feedback. Residual gases and, possibly, adsorbed gas atoms are ionized at the latter stages of the multiplier where electron current densities are highest. If a fraction B of the total number of positive ions bombards a preceding stage and causes γ secondary electrons to be emitted per incident ion, regenerative breakdown occurs when $B\gamma\mu = 1$, where μ is the current amplification between the stages involved. Just below breakdown voltage, dark current due to ion feedback increases more rapidly than photocurrent with increasing voltage. Photon feedback to the cathode occurs when electrons excite fluorescence of the bulb wall and dynode spacers. Electrons which produce the initial ionization or fluorescence may result from field emission or from amplified thermionic emission. Field emission alone may cause high dark currents if strong internal fields are present. Proper distribution of dynode potentials and removal of sharp edges minimize field emission. The in-line multiplier structure is superior to the circular structure with respect to distribution of potentials and isolation of output from input.

A useful figure of merit for comparing dark currents of multiplier phototubes is obtained by expressing the anode dark current in terms of the radiant, or luminous, flux input which would produce an anode photocurrent equal to the dark current. This flux, called "equivalent anode-dark-current input," is simply the anode dark current divided by the anode sensitivity. Equivalent anode-dark-current input plotted as a function of anode sensitivity is a useful multiplier phototube characteristic. It shows a broad minimum extending over a range of sensitivity for which the ratio of photocurrent to dark current is maximum.

If the dark current were constant, it could be eliminated from the output circuit by suitable null methods or by modulating the input flux and amplifying the alternating component of photocurrent. However, the limiting dark current, chiefly amplified thermionic emission, fluctuates as a result of statistical variation in the rate of emission of thermionic electrons from the photocathode. This fluctuation, or shot noise, consists of current pulses having a gaussian distribution of amplitudes. The rms value of shot-noise current measured at the cathode is given by[23] $(2e\,\Delta f I_t)^{1/2}$, where e is the electron charge, Δf is the bandwidth of the circuit in which the noise current is measured, and I_t is the average value of the thermionic emission current.

The noise current is amplified by the current amplification factor μ_n in the multiplier and, because of the random nature of secondary emission, the noise increases by a factor m where $m = [1 + b/(\mu_o - 1)]$ 23, 24, b = 1.5 a constant, μ_o = current amplification per stage at first few dynodes. Thus, at the anode, the rms shot-noise current is given by

$$I_n = \mu_n (2em\ \Delta f\ I_t)^{1/2} \tag{2}$$

I_n is proportional to $(\Delta f)^{1/2}$ up to mid-band frequencies (50-100 megacycles per second) at which the spread in electron transit time begins to broaden the discrete noise pulses resulting from current amplification of individual thermionic electrons emitted at the photocathode.

The amplified shot-noise current may be expressed in terms of radiant or luminous flux input. The flux in this case is modulated in a specified manner to produce an alternating "signal" at the anode. The rms value of the signal I_s produced by a known flux input may then be used to compute the signal-to-noise ratio, I_s/I_n. The bandwidth Δf customarily is specified as one cycle per second. From Eq. (2), signal/noise may be computed for other bandwidths if the dependence of I_s on bandwidth is also known. That value of flux for which $I_s = I_n$ is called "equivalent noise input."

Square-wave rather than sinusoidal modulation of the light input is frequently used because it can be obtained by use of a light chopper — a simple and readily reproducible device. If the "on" and "off" intervals are equal, the amplitude of the fundamental component of the square wave of anode photocurrent is $2I_b/\pi$, where I_b is the anode current which flows during the "on" time of the square wave. Only this component is passed by a filter having a bandwidth of one cycle per second. Thus, the rms signal measured at output of the filter is given by

$$I_s' = (2)^{1/2}\ \frac{I_b}{\pi} = (2)^{1/2}\ \frac{\mu_n S_k F}{\pi} \tag{3}$$

where $I_b = \mu_n S_k F$,

S_k = cathode luminous sensitivity in amperes per lumen or watt

F = flux input.

The test flux F is usually such that $(I_s')^2 >> (I_n')^2$. Under these idealized conditions, the signal-to-noise ratio is

$$\frac{I_s'}{I_n'} = \frac{S_k F}{\pi (em\ I_t)^{1/2}} \tag{4}$$

The equivalent noise input (ENI) is that value of F for which $I_s' = I_n'$, or

$$ENI = \frac{F}{(I_s'/I_n')} = \pi (em I_t)^{1/2}/S_k \tag{5}$$

ENI is quoted in published data on multiplier phototubes and, in some instances, is given as a function of temperature. From this characteristic and Eq. (5) the thermionic emission I_t can be calculated

Because the rms shot-noise current in a bandwidth of one cycle per second (cps) is very small (micro micro amperes for typical tubes) a larger bandwidth (1000 cps) is commonly used. Harmonics of the square wave of anode current which are passed by the filter contribute to the signal I_s. Thus, with a low-pass filter which has an effective bandwidth Δf (equals 1/4RC) of 1000 cps for "white" noise, the contribution from the harmonics of a 90-cps square wave increases the rms value of I_s' by a factor of 1.060. Under these test conditions the measured signal-to-noise ratio is

$$\frac{I_s}{I_n} = \frac{1.060\ S_k\ F}{(1000)^{1/2}\ (em\ I_t)^{1/2}\ \pi} \tag{6}$$

or a factor of $1.060/(1000)^{1/2}$ lower than the value which would be obtained in a bandwidth of one cps as specified in the definition of ENI. Accordingly,

$$ENI = \frac{1.060\ F}{(1000)^{1/2}\ (I_s/I_n)} \tag{7}$$

The signal-to-noise ratio theoretically realizable at the photocathode is degraded if a portion of the photocurrent fails to enter the multiplier. The ratio of effective photocurrent, or that which is amplified and produces anode current, to the total cathode photocurrent is called "collection efficiency." This ratio is difficult to measure directly — for example, by measuring cathode, dynode No. 1, and focus-electrode currents. Secondary emission from the focus electrode and first dynode usually obscures the true distribution between these electrodes.

Relative collection efficiency as a function of position on the cathode can be determined by the use of flying-spot scanning of the entire cathode area. Instantaneous amplitude of anode current may be measured on an oscilloscope for a series of single-line scans. However, secondary emission from the focusing electrode and nonuniform cathode sensitivity or first-dynode emission are not readily distinguishable from variations in collection efficiency.

Average collection efficiency may be computed if the photocurrent I_k which actually enters the multiplier is known. An indirect measure of this photocurrent is obtained from its associated shot noise in a standard test.[25] Constant photocurrent I_k is produced by an unmodulated flux F. I_k may replace I_t in Eq. (6) because both types of emission generate shot noise. Thus, the ratio of rms signal current I_s to the rms noise-in-signal current I_{ns} is

$$\frac{I_s}{I_{ns}} = \frac{1.060\ S_k F}{\pi\ (1000\ me\ S_k F)^{1/2}}$$

$$= 2.67 \times 10^7 \left(\frac{\epsilon S_k' F}{m}\right)^{1/2}$$

under the test conditions previously stated. S_k is the

effective cathode sensitivity in amperes per unit flux or the apparent cathode sensitivity S'_k (measured between cathode and all other electrodes connected together as anode) times the collection efficiency ϵ. In Fig. 9, ϵ is shown as a function of S'_k and signal/noise-in-signal, which is expressed in decibels by the relation $S/N_S = 20 \log I_S/I_{NS}$. As an example, assume cathode sensitivity $S'_k = 65$ microamperes/lumen, $S/N_S = 35$ db when $F = 10^{-7}$ lumen so that $S/N_S - 10 \log F = 105$, and that current amplification of first dynode is 5. From point where the line for $\mu = 5$ intersects 65 microamperes per lumen project a line vertically to the line for $S/N_S - 10 \log F = 105$. Collection efficiency ϵ at this intersection point equal 94%.

The ratio I_S/I_{NS} is of fundamental importance in applications requiring the detection of variations of light level in the presence of relatively high background illumination, and in scintillation counting. In applications involving low contrast, shot noise in the photocurrent associated with background illumination tends to obscure the contrast information. In scintillation-counting applications light pulses of uniform amplitude give rise to anode-current pulses which vary in amplitude as a result of the statistical fluctuation in the number of electrons N_e per pulse which arrive at the first dynode.

Although I_S/I_{NS} is a useful criterion for tubes used in scintillation counting, a more functional test is used occasionally in final evaluation of these tubes. The tube under test is optically coupled to a crystal scintillator such as thallium-activated sodium iodide. Scintillations are excited in the crystal by the action of gamma radiation from a radioisotope such as Cs^{137}, and the resulting current pulses at the anode have a statistical dispersion in amplitude. Pulse-height resolution of the phototube and the crystal scintillator combination is a measure of this dispersion.

To measure pulse-height resolution, a frequency distribution of anode-current pulse amplitudes may be

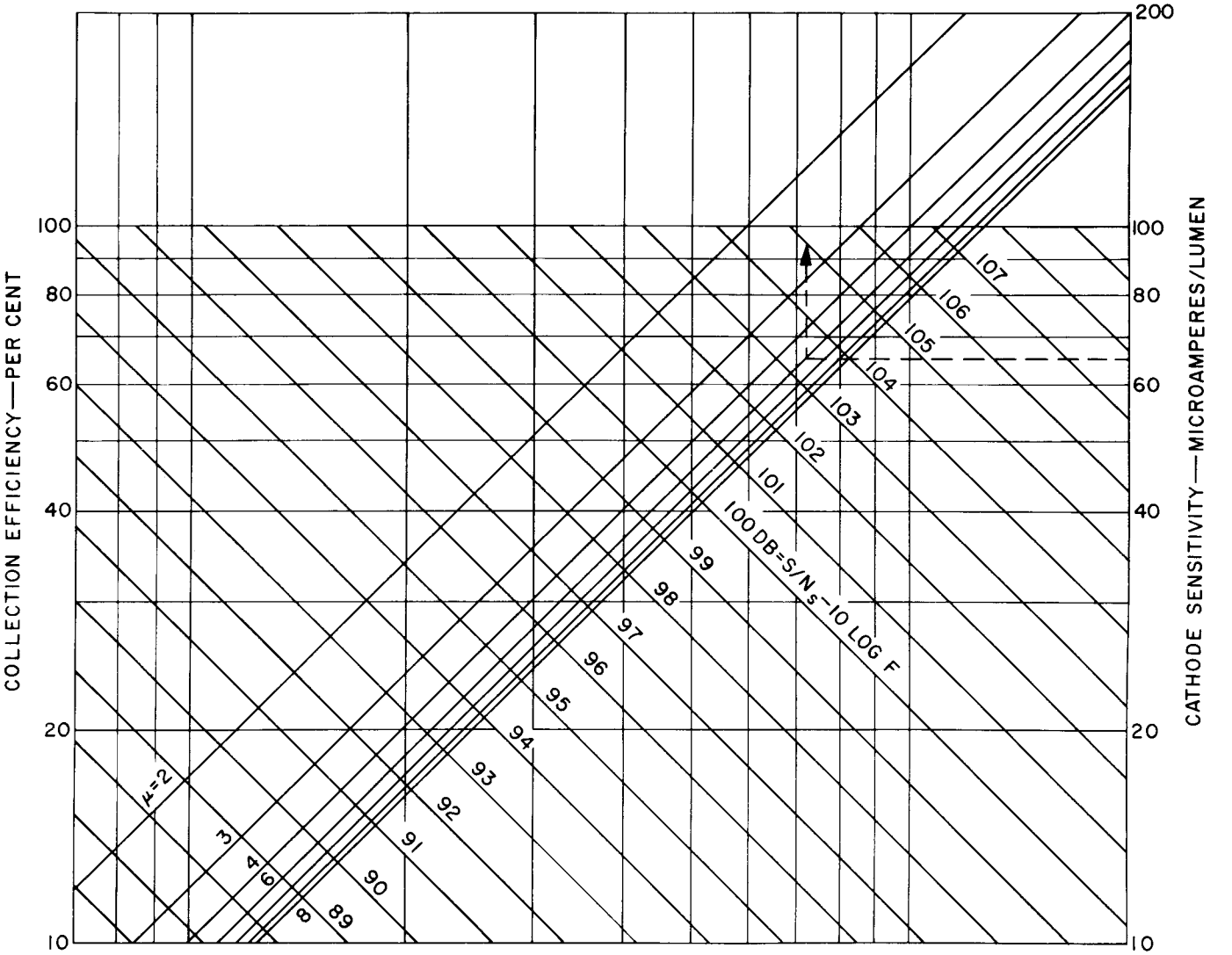


Figure 9. Collection Efficiency as a Function of Measured Cathode Sensitivity, Signal/Noise-in-Signal, and First Dynode Gain, μ_o

plotted as a function of pulse height, as in Fig. 10. The number of pulses n per unit time which have an amplitude between E and E + ΔE may be measured by means of a pulse-height analyzer.[26] The amplitude E is normally obtained by amplifying linearly the pulse voltage developed across the phototube load impedance by the anode current. The mean amplitude E_0 is proportional to the mean flux per scintillation incident on the photocathode. Intensity of the scintillations is proportional to the quantum energy of gamma rays totally absorbed in the crystal. Because of spectral energy distribution of the scintillations is independent of quantum energy of the gamma radiation, the pulse height E_0 is directly proportional to the gamma energy and may be calibrated accordingly for gamma spectrometry.

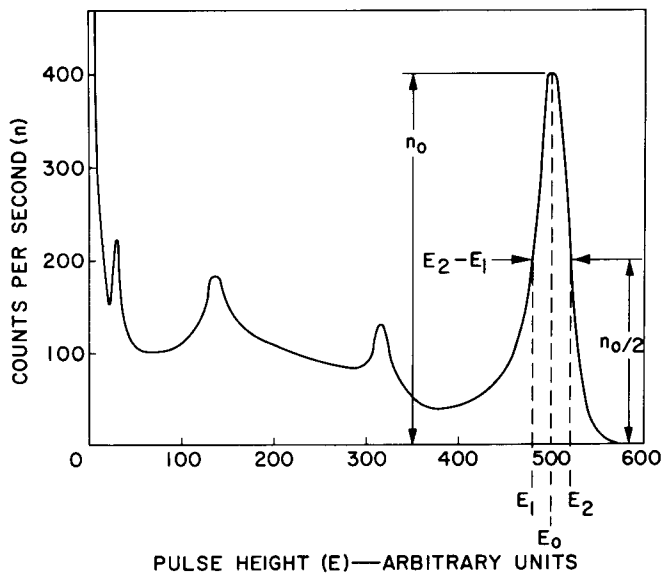


Figure 10. Pulse-height Spectrum of Cs^{137} as Obtained from a Scintillation Counter Using NaI(Tl) Crystal and a Single-Channel Pulse-Height Analyzer. By Definition Pulse-Height Resolution is E_2-E_1/E_o

Pulse-height resolution is a measure of the dispersion introduced by the phototube in gamma spectrometry of the 662,000-electron-volt gamma rays from the radioisotope Cs^{137}. The distribution of pulse amplitudes E is approximately gaussian about a mean value F_0 corresponding to the 662,000-electron-volt spectral line. The width of the curve in Fig. 9 is measured between points E_2 and E_1 at which the count rate n is 50 per cent of the peak count rate n_0. Pulse-height resolution is defined as $(E_2-E_1)/E_0$. Pulse-height resolution is proportional to $(ES'_k)^{-1/2}$ where, in this case, S'_k is the apparent cathode sensitivity to radiation having the spectral energy distribution of the scintillations.

At the low end of the gamma spectrum (Fig. 10), pulses due to dark current in the phototube are indistinguishable from pulses resulting from incident gamma or X-radiation. Hence, the distribution of noise-current pulses is of importance in the design of multiplier phototubes intended for use in low-energy gamma spectrometry.

The multiplier phototube should preserve the time-discrimination as well as the pulse-height resolution capabilities of the scintillator. Many organic and some inorganic scintillators have decay times of the order of 10^{-8} second or less. "Time-discrimination capability" of a multiplier phototube is a measure of the minimum time interval which can separate two individually discernible pulses of anode current resulting from uniform light pulses of negligible duration. The current pulses are considered distinguishable if the amplitude between pulses decreases to 50 per cent of the pulse amplitude.

Time-discrimination capability of the phototube is limited largely by the spread in electron transit time between cathode and first dynode and by the transit time between dynodes. The time required to excite secondary and photoelectric emission appears to be two or three orders of magnitude shorter than the variation in transit time.[27] Differences in transit time between various regions of the cathode and first dynode are now measured by scanning the cathode with light from a high-pressure hydrogen arc, which is capable of producing light pulses with a rise time of 5×10^{-10} second. By means of this technique, the transit-time spread in the cathode-to-first-dynode region has been reduced to about 10^{-10} second over a cathode diameter of 1.5 inches.[28]

Other design tests are performed whenever the application requires special characteristics. Notable among these are stability of sensitivity (required in scintillation counting and critical photoelectric control service), electrical leakage between dynodes, and peak anode current output.

REFERENCES

1. Prescott, C.H., Jr., and M. J. Kelly,"The Caesium-Oxygen-Silver Photoelectric Cell," Bell System Technical Journal, Vol. XI, p. 334, July 1932
2. Sommer, A., "Photoelectric Tubes," 2nd Ed., Methuen and Co., 1951
3. Görlich, P., "Composite Transparent Photoelectric Cathodes," Z. F. Physik, Vol. 101, p. 335, 1936
4. Sommer, A., "New Photoemissive Cathodes of High Sensitivity, "Review of Scientific Instruments, Vol. 26, No. 7, p. 725, July 1955
5. Sommer, A., "Multi-Alkali Photocathodes," IRE Transactions on Nuclear Science, Vol. NS-3, No. 4, p. 8, Nov. 1956
6. Spicer, W. E., "Photoemissive, Photoconductive, and Optical Absorption Studies of Alkali-Antimony Compounds," Physical Review, Vol. 112, No. 1, p. 114, October 1, 1958
7. Engstrom, R. W., et al, "Recent Work on Photoemission and Dark Emission Problems," IRE Transactions on Nuclear Science, Vol. NS-5, No. 3, p. 120, 1958
8. Engstrom, R. W., "Calculation of Radiant Photoelectric Sensitivity from Luminous Sensitivity," RCA Review, Vol. XVI, No. 1, p. 116, March 1955
9. Dekker, A. J., "Secondary Electron Emission," Solid State Physics, Vol. 6, Academic Press, 1958
10. Sommer, A., "Activation of Silver-Magnesium and Copper-Beryllium Dynodes," Journal of Applied

Physics, Vol. 69, p. 598, 1958
11. Bruining, H., Secondary Electron Emission, McGraw-Hill, 1954
12. Code V1
13. Zworykin, V. K., G. A. Morton, and L. Malter, "The Secondary Emission Multiplier — A New Electronic Device," Proc. IRE, Vol. 24, No. 3, p. 351, March 1936
14. Sternglass, E. J., "High-Speed Electron Multiplication by Transmission Secondary Electron Emission," Review of Scientific Instruments, Vol. 26, No. 12, p. 1202, December 1955
15. Kleynen, P. H. J. A., "The Motion of an Electron in Two-Dimensional Electrostatic Fields," Philips Technical Review, Vol. 2, p. 338, 1937
16. Almas, G. A., G. Diemer, and H. Groendijk, "A Rubber Membrane Model for Tracing Electron Paths in Space Charge Fields," Philips Technical Review, Vol. 14, No. 11, p. 336, May 1953
17. Mayo, B. H., "A Rubber-Membrane Model for Axially-Symmetric Electric Fields, "British Journal of Applied Physics, Vol. 6, No. 4, p. 141, April 1955
18. Allen, D. N, R. V. Southwell, and G. Vaisey, Relaxation Methods Applied to Engineering Problems. Part XI. Problems Governed by the "Quasi-Plane-Potential Equation," Proc. Roy. Soc. (A), Vol. 183, p. 258, 1954-55
19. Zworykin, V. K., and G. A. Morton, Television, 2nd Ed., Wiley and Sons, 1954
20. DePackh, D. C., "A Resistor Network for the Approximate Solution of the Laplace Equation," Review of Scientific Instruments, Vol. 18, No. 10, p. 798, October 1947
21. Salinger, H., "Tracing Electron Paths in Electric Fields," Electronics, Vol. 10, No. 10, p. 50, October 1937
22. Engstrom, R.W., "Multiplier Phototube Characteristics: Application to Low Light Levels, "Journal of the Optical Society of America, Vol. 37, No. 6, p. 420, June 1947
23. Van der Ziel, A., Noise, Prentice-Hall, 1954
24. Morton, G.A., "Photomultipliers for Scintillation Counting," RCA Review, Vol. X, No. 4, p. 525, December 1949
25. Engstrom, R.W., R.G. Stoudenheimer, and A.M. Glover, "Production Testing of Multiplier Phototubes Designed for Scintillation Counter Applications," Nucleonics, Vol. 10, No. 4, p. 58, April 1952
26. Francis, J.E., et al, "Single-Channel Analyzer," Review of Scientific Instruments, Vol. 22, p. 133, March 1951
27. Code V2
28. Morton, G.A., et al, "Design of Photomultipliers for the Sub-Millimicrosecond Region, "IRE Transactions on Nuclear Science, Vol. NS-5, No. 3, p. 98, 1958

Camera Tube Design and Processing

B. H. Vine

Lancaster

A camera tube is an electron tube that converts an optical image into an electrical signal by a scanning process. This discussion is limited to the storage type of camera tube* in which the light from a given picture element is utilized during an appreciable fraction of the frame time. In all existing storage-type tubes, the light produces an electrical charge which is stored on the surface of a target during the interval between scans.

TARGETS

Diagrams of the targets of the image orthicon, vidicon, orthicon, and iconoscope are shown in Fig. 1. The conducting plate b (for backplate electrode) is incorporated in each target in order to provide every element of the storage surfaces with a definite capacitance. Between the storage surface and the backplate is the dielectric d.

In the image orthicon, the storage surface is a thin glass membrane, the dielectric is vacuum, and the backplate takes the form of a mesh through which the photoelectrons may pass to the storage surface. The membrane emits more than one secondary electron per incident photoelectron and consequently is driven positive in illuminated areas. In addition to its function as backplate, the mesh serves as a collector for the secondary electrons. The membrane is made of such thickness and conductivity that no large potential difference builds up between its two surfaces, yet no appreciable lateral conduction, which would reduce resolution, occurs between picture elements.

In the vidicon, the dielectric layer is a photoconductive solid. The charge distribution is built up by photoconduction from the backplate electrode, which is transparent and is ordinarily called the signal electrode because the signal is taken from it.

In the iconoscope and orthicon, the photoemitter is on the storage surface. Since useful photoemitters have lateral conductivity, the photoemissive surface is laid down as either an array or as a random mosaic of islands on the dielectric, which is a good insulator.

Deposition of Signal Charge on the Target

In all of the camera tube types considered here, each scan of the electron beam brings the storage surface to, or nearly to, a fixed potential. For types employing low-velocity scanning this fixed potential is close to that of the thermionic cathode. For types employing high-velocity scanning this potential is close to that of the collector of the secondary electron produced by the scanning beam. In each type of tube (except the iconoscope) there is a potential, different from that determined by the beam, toward which light drives the storage surface. The difference between this potential and that of the storage surface will be called the driving potential difference.

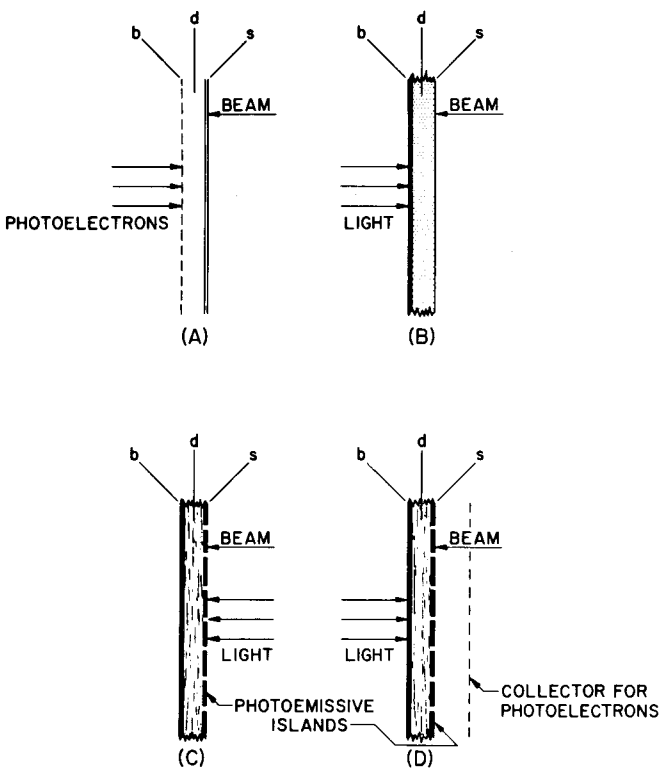


Figure 1. Camera Tube Targets: (A) Image Orthicon or Isocon; (B) Vidicons; (C) Iconoscope; (D) Orthicon (b = backplate electrode; d = dielectric; s = storage surface)

*For definitions of terms used in this chapter see Ref. 1. For descriptions of the operating principle of the various types of camera tubes see Refs. 2, 3, and 4 or the RCA tube bulletins.

In the image orthicon, the mesh is maintained at a potential about 2 volts above that of the thermionic cathode. When the storage surface approaches mesh potential, the driving potential difference becomes small, most of the photoelectron-produced secondary electrons fall back onto the target, and the amount of charge on the glass membrane approaches saturation.

In the vidicon, the signal electrode is typically maintained at a voltage of 40 volts above the potential of the thermionic cathode to provide driving potential difference. Ordinarily, the signals used in the vidicon produce a change in storage surface potential of only a few volts. Thus, the stored charge does not approach saturation.

The orthicon as originally built had a large driving-potential difference, approximately equal to the wall-electrode-to-thermionic-cathode potential difference, so that saturation was never approached. The subsequent introduction of a mesh collector for photoelectrons at a few volts above the potential of the thermionic cathode allows signal saturation to occur just as it does in the image orthicon.

Scanning, Signal Extraction, and Erasure

Ideally, the scanning beam deposits charge on the target storage surface equal and opposite to that previously deposited by light, thereby bringing the storage surface back to a fixed, uniform potential. The signal may be extracted in various ways:

In all tubes, the signal might be obtained as backplate electrode current.
The signal level, however, may be so low at the target that it would not be adequately above the noise level of the amplifier circuit. Such is the case with image orthicons. To obtain an adequate signal in image orthicons, it is necessary to pass the returned electrons that were not required to discharge the signal charge on the target on to an electron multiplier from which a high-level, inverted signal can be obtained.

In the isocon, the back-scattered electrons (whose number is about proportional to the number of electrons deposited by the scanning beam) are separated from the specularly reflected electrons and enter a multiplier to provide a signal of normal polarity.

In the iconoscope and vidicon, the target signal is made high enough so that it can be taken out at the backplate.

The erasure of signal charge from the target by the beam is never quite complete, so that there is always residual signal on the storage surface carried over from one frame to the next. This effect has been given the name capacitance lag.

High-Velocity Scanning versus Low-Velocity Scanning

When high-velocity scanning is used, the beam is not easily deflected and beam bending is negligible; beam-landing errors are not a problem (in fact, the entire beam always lands); capacitance lag tends to be small because the excess of beam current used gives a more complete erasure, and also, because redistribution electrons tend to erase the signal already present on the target.

On the other hand, with low-velocity scanning there is no problem with secondary electrons falling back on the target with consequent reduction or modification of the charge pattern. In addition, low velocity scanning is less sensitive to contaminated areas on the target (these areas will show up prominently with high-velocity scanning when their secondary-emission ratio is different from that of the remaining areas).

TARGET DESIGN

Target Size

The size of the target more or less determines the size of the camera tube and the power required for deflecting and focusing the beams. Because increasing the target size increases the tube size and power requirements, it is generally desirable to keep the target as small as possible. On the other hand, high resolution is more readily obtained with a large target. Also, high target capacitance, which may be required for a high signal-to-noise ratio, is more readily obtained when the target is large. Furthermore, if the target size involves a dimension which must be kept small compared to the element size, such as that of the mesh pitch or the target thickness, then a large target is simpler to fabricate.

It is conceivable that in the future the available current density at the thermionic cathode may set the target size. In tubes employing low-velocity scanning, where the beam electrons are decelerated to approximately cathode potential, the scanning-spot current density cannot be greater than the thermionic-cathode current density. Where a large stored charge with high resolution and fast scanning are required, the scanning spot current density might have to be comparable to thermionic-cathode current densities.

Target Surface-Potential Swing

The deposition of charge on the target by the beam is controlled by the potential of the target surface. The change in surface potential during the passage of the beam over a given element will be called surface potential swing or potential swing. It is necessary to choose a compromise value of potential swing. Too small a potential swing leads to excessive capacitance lag, while too large a potential swing may lead to beam bending and related effects (for low-velocity scanning), to lateral leakage problems, or to signal saturation if the swing is comparable with the driving potential difference. In the image orthicon, too small a potential swing gives low percentage beam modulation and, consequently, excessive noise; too large a potential swing will yield beam electron secondary-emission ratios approaching unity with, again, low percentage beam modulation.

Target Charge

The amount of charge stored on the target affects the

signal-to-noise ratio which may be obtained. For example, see the text headed Charge Required on an Image Orthicon Target.

Some Target Relationships

Let m = number of elements in raster

Let t = active scanning time for these m elements

Then B = $\frac{m}{2t}$ is the bandwidth required to transmit $\frac{m}{t}$ elements per unit time.

Let A = raster area on target

Let d = storage-dielectric thickness

Let ϵ_0 = permittivity of vacuum

Let ϵ_r = relative permittivity of dielectric

Then C = $A\,\epsilon_0\epsilon_r/d$ = total target capacitance in rationalized MKS units.

Let c = $\frac{C}{m}$, the element capacitance

Consider a target uniformly charged with charge Δq per element, then, the total charge $\Delta Q = m\Delta q$. The instantaneous target current is $i_t = \Delta Q/t$. The target potential swing ΔV is $\Delta V = \Delta Q/C = \Delta q/c$.

Calculation of Capacitance Lag

Consider low-velocity scanning. Let the distribution of axial electron energies be given in integrated form, that is, the amount of current which lands on a surface of potential V as a function of V with constant incident beam current. A typical curve of I(V) is shown in Fig. 2. Call this the beam deposition characteristic.

Suppose, for convenience, that the scanning spot is rectangular, has uniform density, just covers one element for the time t/m and is moved stepwise from element to element. The current I(V) decreases the charge on the element, hence $dq/dt = -i(V)$ where q is the charge on the element. Since $q = cV$, substitution gives: $c(dV/dt) = -I(V)$. Integration gives:

$$\int_{V_0}^{V_1} \frac{dV}{I(V)} = -\frac{1}{c}\int_0^{t/m} dt = -\frac{t}{mc} = -\frac{t}{C}$$

If I(V) is given, 1/I(V) as a function of V may be obtained as in Fig. 3. Then $\int dV/I$ may be determined as in Fig. 4. If I(V) is given as an experimentally determined curve, then graphical or numerical integration must be used to obtain the curve of Fig. 4.

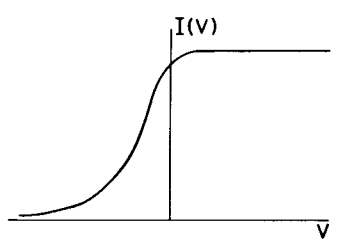


Figure 2. Beam Deposition Characteristic

To interpret Fig. 4, note that the dimensions of $\int dV/I$ are those of time divided by capacitance It is seen that in the preceding equation the change in the value of this integral during scanning is equal to the element time divided by the element capacitance (or, alternatively, the active scanning time divided by the total active target capacitance).

Let the potential of a given element of storage surface be V_0 just before scanning, and let the capacitance of this storage element be $_c$. The potential V_0 establishes an initial point on the curve of Fig. 4. If the scan spot remains on the element for a time t/m, then the value of the ordinate of the curve decreases (negative sign in equation) by an amount t/mc. The new ordinate determines the final abscissa V_1, which is the potential of the storage surface after scanning. If, during the interval between scans, a charge Δq is added to the storage element, then the potential will increase by an amount $\Delta q/c$ to obtain a new value for the potential which then yields a new initial point for the following scan. The preceding general process may be repeated over and over.

In the steady state in which the same charge Δq is repeatedly stored between each scan and the next, the steady state values of initial and final potential may be

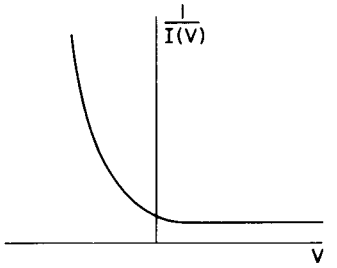


Figure 3. Reciprocal of Beam Deposition Characteristic

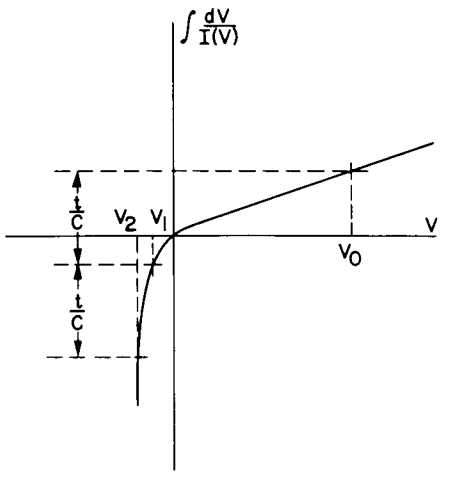


Figure 4. The Time vs. Potential Function

found as the two points on the curve which are separated by an abscissa distance $\Delta V = \Delta q/c = It/c$ and by an ordinate distance t/mc. The character of the curve of Fig. 4 insures that there will either be one such pair of points to satisfy this condition or no points which do. The latter case corresponds to insufficient beam.

To determine the capacitance lag in the image orthicon, assume that the potential V_0 and V_1 are the steady state values of initial and final potentials. If, after a given scan, no signal is added, then the storage surface potential remains at V_1 and this becomes the initial potential for the next scan during which the ordinate changes again by t/mc and the surface potential goes to V_2. Thus the scanning process deposits a charge $(V_2 - V_1)/C$ on the target and a signal corresponding to this charge appears at the tube output even though no charge was stored. This is lag signal. This process may be repeated with V_2 as the new initial potential and another scan. The second lag signal will be smaller than the first and so on.

A slightly more elaborate problem is one in which there is a nonzero value of Δq in the dark such as due to dark current in the vidicon, or such as would be obtained by adding a uniform illumination to the image projected on the image orthicon faceplate. The latter will be called bias light. This type of problem is handled by the general method.

Another problem which may be solved by this technique is that of resolving an observed vidicon lag into its capacitance and photoconductive components, which can be done if the capacitance and beam-deposition characteristic are known.

It should be noted that the assumption of stepwise motion of the scanning spot was made for simplicity. The same result is obtained for continuous motion.

The curve corresponding to Fig. 4 may be determined analytically for a Maxwellian distribution of electron energies defined by a temperature T, and a total current I_0. Such a curve is given in Fig. 5.

Analysis of various cases with the curve of Fig. 5 indicates that the potential swing of the first residual

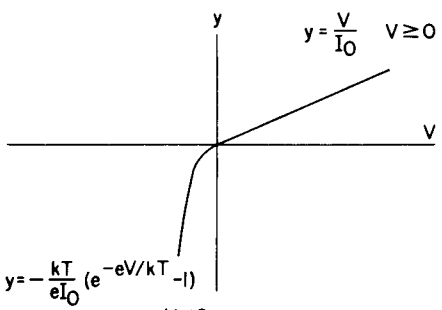


Figure 5. The Time vs. Potential Function for Maxwellian Distribution

signal will correspond to one to two times the equivalent voltage of the distribution. That is, for an 1100 K distribution, the first residual signal will be 0.1- to 0.2-volt potential swing. Thus, to minimize capacitance lag, the signal-potential swing should be made large compared to a voltage of 0.1 or 0.2 volt.

Charge Required on an Image Orthicon Target

Even in the steadiest, most uniform light, photons fall at random. The emission of an electron by a photon is also a random event. In such a case, the number of photoelectrons per element per scanning time required to give a signal-to-noise ratio of R is R^2, according to statistical theory. If this effect were the only source of noise, the total charge required to be stored on the target per scanning time would be meR^2A where e is the electronic charge and A is the charge amplification factor at the target. When all secondary electrons are collected (below the knee of the transfer characteristic) the factor A is δ-1 where δ is the secondary emission ratio. In view of other sources of noise, a larger charge $mcR^2A\Sigma$ is needed. The factor Σ is always larger than unity and is defined in the next section.

Noise in the Image Orthicon

Consider the photons falling on one picture element during the storage interval. These photons arrive individually and collectively at random. If the average number is n, then according to statistical theory, the rms deviation from the average will be $\sqrt{n}$. The deviations constitute noise. The signal-to-noise ratio is then $\sqrt{n}$. The signal-to-noise ratio of an actual tube will be less than this and may be shown to be $[(E\eta/\Sigma)n]^{1/2}$, where E is the quantum efficiency of the photocathode, η is the mesh transmission factor, and Σ is:

$$\Sigma = 1 + \frac{\delta}{(\delta-1)^2} + \frac{1}{m(\delta-1)} + \frac{1-m}{m\delta_1(\delta-1)}$$

where δ is the secondary-emission ratio for primary electrons from the photocathode incident on the target from the photocathode, δ_1 is the secondary emission ratio for return beam electrons incident as primaries on the first dynode, and m is the beam modulation. In this expression for Σ, the first term (unity) comes from statistical variations in the photoelectrons (primaries) incident on the target, the second term from statistical variations in the number of secondary electrons per incident primary electron, the third term from shot noise in the beam, and the fourth term from statistical variations in the number of secondaries per primary electron at the first dynode. Another form for the expression for Σ is that the ratio of the flux actually used on a tube L to the ideal luminous flux whose statistical variation alone would give the same signal-to-noise ratio L (ideal) is $L/L_{(ideal)} = \Sigma/E\eta$. This figure is analogous to the "noise figure" used in radio, except that photon shot noise is used as the basis of comparison rather than thermal noise. The figure $L/L_{(ideal)}$ is just the reciprocal of a term called detective quantum efficienty[5]. The expression for Σ is derived on the assumption that all secondary electrons are collected by the mesh, i. e., operation is well below the knee of the transfer characteristic.

Video Amplifier Noise

The low-level output of a camera tube is coupled to the grid of the first video amplifier stage (see Fig. 6). A vidicon is used as an example. The low-level parts, including the signal electrode in the vidicon, the leads, and the grid in the amplifier tube, inevitably have capacitance to ground which tends to shunt the high-frequency components of the signal to ground. The later stages of amplification are made with increased high-frequency gain to compensate for this effect. The signal-to-noise ratio due to thermal noise in the load resistance and shot noise in the tube is found to be

$$\frac{Ri_S}{\{4kTB[R + R_t(1 + \frac{4}{3}\pi^2R^2C^2B^2)]\}^{1/2}}$$

where R is the load resistance, C is the total capacitance of the low-level parts to ground, R_t is the equivalent noise resistance of the first amplifier tube, B is the bandwidth, and i_s is the signal current. Note that the signal-to-noise ratio is proportional to the signal current.

This expression is not fully satisfactory because it indicates an optimum value for R indefinitely large, while in practice a fairly definite optimum value of R is obtained. The values of R obtained are such that the dimensionless product RCB is about 4 when B is taken as the bandwidth in cycles necessary to transmit the useful signal information. C is usually about 30 micromicrofarads. Perhaps the expression is deficient because it does not include the noise originating in the later stages; such noise is probably significant at the high end of the frequency band.

Capacitance Effects in Wide-Spaced Image Orthicons

Image orthicons in which the target-to-mesh spacing is considerably greater than the element size are called wide-spaced. When such tubes are operated at a light level above that at which the knee would be expected on the basis of the parallel-plate capacitance, they exhibit even more edge emphasis than do conventional tubes. This occurs because the edges of an illuminated area have more capacitance than the remainder of the illuminated area. Another effect that occurs in wide-spaced tubes is that the signal rises with illumination even above the calculated knee. This effect occurs because the area which has just been scanned is at cathode potential and, therefore, the area which is about to be scanned has appreciable capacitance with respect to the freshly scanned area. Consequently, the area about to be scanned has its potential lowered by the proximity of the freshly scanned area. It follows that the area to be scanned can be given extra charge by the excess of light before the area is scanned.

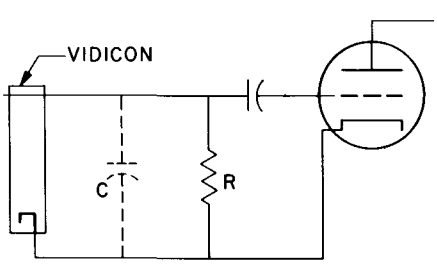


Figure 6. Coupling Impedance between the Vidicon and Its First Preamplifier Stage

ELECTRON OPTICS

This section covers only focus by means of a long magnetic field. Other methods involving electrostatic lenses and short magnetic lenses are possible. With the small currents used in camera tubes, space-charge effects are not significant except near the focal point of a beam decelerated to near cathode potential. Electrons emitted from any type of cathode have a distribution of initial velocities. Much of the design effort in electron optics is concerned with obtaining a reasonable focus in spite of these different initial velocities. Much of the answer to this problem is to reduce the significance of the initial velocities by making the potential of the space through which the electrons travel as high as possible.

Focusing

Two simplified and contrasting illustrative examples follow:

1. Suppose that electrons are emitted with total initial energies between zero and V_0 volts from a cathode into uniform and parallel magnetic and electric fields (see Fig. 7). This example is an approximation to the image section of an image orthicon. The projection of the path of any electron on a plane perpendicular to the fields will be a circle. The electrons which start with zero axial component of initial velocity will come to a point focus at each of the distances

$$L = 10.6\frac{N\sqrt{V}}{B} \text{ centimeters}$$

when B is the magnetic field in gausses, V is the potential at L with respect to the cathode in volts, and N is any integer. It has been shown[6] that around each of these points the circle of confusion of all electrons has a radius $L(V_0/V)$. This focus is not necessarily the "best" focus, but is chosen for comparison with the next example.

2. Suppose electrons are injected all with the same total energy corresponding to V volts into a uniform magnetic field B and a zero electric field (constant potential V). (See Fig. 8.) The distance required to bring electrons injected at an angle α from the field direction are brought to a focus at the distance

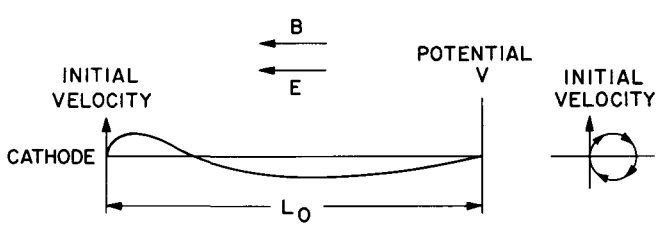


Figure 7. Electron Path in Parallel and Uniform Magnetic and Electric Fields (B = magnetic field; E = electric field)

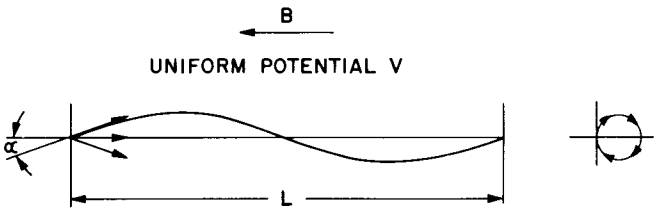


Figure 8. Electron Path in Uniform Magnetic Field, and Constant Electric Potential (B = magnetic field)

$$L = 21.2 \frac{N\sqrt{V}}{B} \cos \alpha \text{ centimeters}$$

Because L varies with the angle of injection α, no perfect focus is possible if electrons enter at all angles between zero and some limiting value of α which will be called α_0. When L is chosen for focus of paraxial rays, the radius of the circle of confusion for small values of α_0 is

$$L \frac{\alpha_0^3}{2} \cong \frac{L}{2}\left(\frac{V_0}{V}\right)^{3/2}$$

where V_0 is the maximum lateral component of energy, $V \sin^2 \alpha_0$.

This example corresponds to the use of a gun in which the electrons leave the cathode without initial thermal velocities, but in which some of the electrons have been given lateral energy by the electric or magnetic fields within the gun, and have lost corresponding axial energy in gaining their lateral component of energy. On the other hand, if the lateral component of energy had been due to initial thermal velocities, these thermal velocities would have produced no loss of axial energy and the focus would have been perfect. Actually no such perfect focus is possible because an axial distribution of thermal energies will always be present. It is important to distinguish between lateral energies due to electron optics and those due to thermal effects.

The preceding examples show that it is desirable to have the space potentials high and the distances short to obtain a focus spot of the smallest size.

Deflection Aberrations

Deflection defocusing (field curvature) is difficult to avoid with flat-target tubes, except by making the deflection angle small. Even this solution requires compromise because excessive beam-path length tends to increase the on-axis aberrations. Periodic parabolic wave form variation of the focus voltage at horizontal and vertical frequencies can be used to overcome the problem of deflection defocusing. There are additional deflection aberrations, such as astigmatism, which cannot be satisfactorily handled by focus voltage variation.

Beam-Landing Errors

When a low-velocity scanning beam is used, it must approach the target with the same axial energy for each point on the raster in order to drive the storage surface to a uniform potential. Lack of such uniformity is called beam-landing error (or porthole) and produces lack of uniformity of signal output.

This lack of uniformity arises because the signal in the vidicon or in the image orthicon (above the knee of the transfer characteristic) varies with the driving potential difference at the beginning of charge storage. The driving potential difference will not be uniform if the storage surface is not driven to a uniform potential.

The beam landing error may be calculated as follows: In the vidicon a fine mesh is mounted at the end of the wall electrode to provide a boundary between the nearly uniform potential space within the wall electrode and the uniform decelerating field region between the mesh and the target. The wall and mesh are called grid No. 3 and grid No. 4 respectively in the 7038 vidicon, and are at the same potential in this tube. The distance between the mesh and the target is small enough that the effect of the magnetic field on the electron while it passes through this region may be neglected. If the electron beam, when deflected, passes through the mesh at an angle other than perpendicular, the electrons do not have as much axial energy as when the beam is undeflected. The target surface is not driven to as low a potential at points away from the center of the raster as at the center. If the angle at which the beam passes through the mesh is α, measured from the normal, then the beam electrons lack $V\sin^2\alpha$ volts of axial energy compared to undeflected electrons, where V is the mesh potential above that of the thermionic cathode. This is the beam landing error, and it is proportional to the mesh potential. It appears that even when no mesh is used, beam-landing errors tend to be greater at higher wall potentials. However, it has always proved possible, through detailed study of beam paths, to make α everywhere small enough so that beam-landing errors are insignificant. If significant beam-landing errors are unavoidable, then a compromise choice of mesh potential V may have to be made between one that will produce low beam-landing errors with low V and one that will produce high resolution at high values of V. (See section on Focusing.)

Limitations on Beam-Current Density

Langmuir and others have shown[7, 8] that the maximum possible current density j at any cross-section of an electron beam is

$$j = j_0\left(1 + \frac{eV}{kT}\right)\sin^2\theta \qquad V \geq 0$$

where j_0 is the current density at the cathode, V is the potential of the point under consideration with respect to the cathode, and θ is the maximum angle the path of any electron makes with beam axis.

In the case of low-velocity scanning, the critical point will ordinarily be directly in front of the target where maximum current density is desired (so that the scan spot may be as small as possible) and where the potential V is approximately zero. It is probable that electrons strike the target at all angles up to grazing incidence, hence, at this point $\sin^2\theta$ is approximately unity. In this case, the maximum current density in

the spot is equal to the cathode-current density. (For retarding potentials, $V \leq 0$, the equation is

$$j = j_0 \, e^{eV/kT} \sin^2\theta.$$

When high-velocity scanning is used, eV/kT is a large number and, therefore, there is no definite limitation to cathode-current density. However, θ usually is small.

Approximate calculations obtained with the equations developed for kinescopes[9, 10, 11, 12] indicate that the focus in camera tubes such as the vidicon is not limited by space charge in the beam. However, equations really suitable for calculating space charge effects with deceleration of the beam electrode to zero potential have not been developed.

In the vidicon, it appears that the secondary electrons emitted from the mesh by the beam electrons move away from the beam slowly because the magnetic field holds them in the vicinity of the beam and because the axial electrostatic fields required to sweep them out are small. The "beam splitting" observed in the vidicon when high beam currents are used is believed to be due to the effect of the space charge of such secondary electrons on the beam. "Beam splitting" can be reduced by operating the mesh a few volts above wall potential in order to return secondary electrons quickly to the mesh.

MISCELLANEOUS SPURIOUS EFFECTS

Sticking and Burn

In general, target phenomena such as "sticking" and "burn" seem to be the same as the phenomena referred to in literature on dielectrics as anomolous charging currents, or dielectric relaxation. These terms refer to the fact that the electrical properties of a capacitor with a given material as dielectric cannot be described in terms of one ideal capacitance in parallel with an ideal resistance.

It has been suggested that burn in the image orthicon is due to the migration of cesium on the surface of the target. This explanation has never been fully proved or disproved.

Vertical Tilt

In both the image orthicon and the vidicon, it is often observed that the signal decreases more or less linearly from the beginning to the end of each vertical field. The decrease is typically about 15 per cent, per unit. It is believed that the effect is due to combination of nonperpendicular scanning and the tendency of the beam to be deflected vertically downward by the positively charged unscanned area below the beam.

Flicker

A certain area or combination of areas, usually including the upper left corner of the raster on the monitor, may give alternate high and low signals ("flicker") on successive fields. There may be more than one type of flicker involved but no satisfactory criteria for classification of this phenomenon have as yet been developed.

Flicker is observed in both the image orthicon and the vidicon. It occurs most readily when the beam is in sharp focus. In vidicons, it occurs when the surface potential swing is large, that is, at high signal levels, and only in tubes with low target capacitance. It occurs in both noninterlaced and interlaced systems. In interlaced systems, correction of improper interlace has at times successfully eliminated flicker. Flicker is strongly affected by beam alignment. It has been found that vidicon flicker can be eliminated by reducing the grid-No.-2 potential.

Waterfall Effect

"Waterfall effect" is a rippling which occurs under conditions similar to those which cause flicker.

Closely related to waterfall effect is "edge crawling" (found in the vidicon only) in which the first and last horizontal lines in each field are bright but not uniform and the nonuniformities change from scanning cycle to cycle. The reason that these lines are bright is that the region of the target outside the raster comes to full signal-electrode potential and, therefore, causes the beam approaching the outer lines to be deflected out of the regular raster to this high-potential region which yields a large signal. This bright edge could be eliminated by limiting the area of the signal electrode to just the scanned raster area. However, in many applications, the vertical blanking is made wider on the picture tube than on the camera tube. This situation removes the lines which show edge crawling.

PERFORMANCE CHARACTERISTICS

Spectral-Response Characteristic

The relative spectral-response characteristic of an image orthicon is just the spectral-response characteristic of the photocathode. The image orthicon may be tested as a diode phototube using the target as the anode.

The vidicon must be operated as a camera tube to obtain significant test results of any kind. Because the vidicon photosurface is nonlinear in its response to light, it is necessary to illuminate the raster uniformly so as to have a known power density. Because of the nonlinearily, the spectral-response curves obtained at constant radial power input have a different shape from the curves obtained at constant output signal. The sensitivity (responsivity) at constant output signal has been chosen for publication.

Transfer Characteristic

To obtain the transfer characteristic of a vidicon, the raster area is uniformly illuminated. The average signal output is then read on a dc meter as a function of the illumination. This method could be used for the image orthicon, but ordinarily, no attempt is made to obtain a carefully defined transfer characteristic for the image orthicon because redistribution phenomona, which are important in determining picture tonal rendition, are not describable in terms of a transfer characteristic.

Resolution

The resolution of a camera tube is ordinarily specified by giving a complete amplitude-response characteristic. When a sine-wave amplitude-response characteristic is available, an equivalent rectangular pass band with quite useful properties may be determined[13]. This band is specified by its maximum television line number N_e. The N_e figure is ordinarily only a fraction, say 1/3, of the limiting resolution observable on a test set monitor.

Noise

Noise is ordinarily specified in terms of the signal-to-noise ratio. The black-to-white signal difference is taken as the signal; the noise is the rms noise within a stipulated bandwidth. No industry standard is yet available for defining the proper noise bandwidth. A further refinement of noise specification might be to specify the distribution of noise within the bandwidth (the noise spectrum).

Persistence Characteristic

The persistence characteristic is normally defined by measuring the decay of signal after a highlight level of illumination on the sensitive surface is replaced by darkness.

Sensitivity

The term "sensitivity" has several different meanings when applied to camera tubes or camera tube systems. The term is used with the phototube meaning in which sensitivity is defined as the ratio of the output current to the input flux (either radiant or luminous). This ratio has also been called responsivity.

The term sensitivity may also refer to the illumination required to bring the tube to some design operating level such as the knee of the image orthicon. In this sense an image orthicon with a large target-to-mesh spacing is more sensitive than a close-spaced one; that is, less illumination is required to reach the knee of the wide-spaced tube.

In addition to the above, there is a figure of merit in which the signal-to-noise ratio is included. This is called detective quantum efficiency (DQE)[5, 14]. This figure is the ratio of the photon flux, whose inherent random variations alone would yield the observed signal-to-noise ratio, to the actual flux of photons required. The DQE is always less than unity and always less than the quantum efficiency of the primary photoelectric process.

PROCESSING OF CAMERA TUBES*

Common Problems

Blemishes on the target or other surface which is at a focus plane must be avoided. It is customary to use specially constructed rooms and cabinets to avoid deposition of atmospheric dust and lint. In addition, tube materials may produce particles which may give spots. These particles include metal splash from spot welding, chips from insulator ceramics, particles produced when bulb spacers enter the tube, particles which come out of the evaporator or getter assemblies, glass particles from drop-seal cullet or from blowing open a manifold tubulation.

Emission of gas or vapor from one tube part may adversely affect the properties of another. This effect may occur either during processing of the tube or during its operation. Secondary-emission ratio, photosensitivity, and thermionic emission from oxide cathodes are all quite sensitive to contamination. In addition, the sticking in glass image-orthicon targets is quite sensitive to contaminants.

Ion spot in the image orthicon is caused by gas in the tube. The beam produces positive ions which collect at the center of the target. Their charge yields a spurious signal. Careful treatment and handling of tube parts, and thorough exhaust keeps ion spot under control. Tubes which have a field mesh are normally operated in a manner such that there is no field tending to concentrate ions at the center of the target and, thus, do not show ion spot.

Materials

Borosilicate glasses, such as Corning 7052 and 7056, have been used for camera-tube bulbs for several reasons. Weathering, which may affect image-orthicon photocathode sensitivity and uniformity, is less of a problem with these glasses than with other glasses and there is reduced risk of cracking a bulb by thermal shock. This reduced risk is important because camera tubes represent substantial investments in labor and material.

It is desirable that a minimum amount of cesium be generated during activation of the image-orthicon photocathode because the excess cesium causes electrical leakage across the target and between the electrodes, and because the most stable and uniform photocathodes are those made with a minimum of cesium. Tube materials for the image orthicon are selected for minimum absorption or adsorption of cesium. Parts made of an 80 per cent nickel 20 per cent chromium alloy are used in preference to stainless steel because the nickel-chromium alloy yields less oxide during sealing of the glass envelope parts. An evaporated-aluminum film has replaced fired-on platinum because aluminum absorbs less cesium. Lead-glass parts are avoided because they take up cesium. Ceramics vary in the amount of cesium they absorb.

Mesh Fabrication

The mesh used in image orthicons and vidicons is electroformed on a glass master. To fabricate a glass master, a polished glass plate is first coated with resist and then ruled, by means of ruling engine, first in one direction and then at right angles to the original directions. The tool used is a diamond in the form of a

*See also chapter on Photoconductors for Camera Tubes.

truncated wedge with the flat cutting surface of a preassigned width to fix the width of the groove cut. The flat surface of the tool rides on the glass surface and does not cut the glass. The resist must be of such quality that it neither chips nor leaves burrs when the tool cuts across a previously ruled groove. After ruling, the plate is etched to a suitable depth and the resist is removed.

The mesh is made from this master by coating the master with a very thin layer of wax to facilitate separation of the mesh and then sputtering the master from a gold-palladum-alloy cathode. The metal-coated master is placed in water and its surface is then rubbed with a pad to remove all metal except that in the grooves. The sputtered deposit must not be so hard that it will not rub off nor so soft that the material removed forms into balls. The hardness of the sputtered deposits is controlled by the conditions in the sputtering operation and the wax parting layer.

The mesh is formed by electoplating the desired amount of metal, either copper or nickel, on the sputtered deposit remaining in the grooves. The mesh is then rinsed, removed from the master, and dried.

The final mesh may be mounted on a support ring having a suitable coefficient of expansion, and the assembly then hydrogen-fired to stretch the mesh. In the case of nickel mesh, the mesh may be placed under mechanical tension and then mounted on a ring to eliminate the need for firing. This method has the advantage of avoiding the annealing of the mesh and the consequent reduction in maximum tension which may be developed in the mesh.

Glass Target Fabrication

The glass membrane of the target is made by heating the closed end of a glass tube in a torch to a suitable softness, and then blowing a large glass bubble about 0.1 to 0.2 mil thick. A portion of this bubble is cut out and laid over the ring on which it is to be mounted. The assembly is heated until the glass softens, adheres to the ring, and is stretched flat by surface tension. The assembly is then cooled at a suitable rate so that the glass membrane ends up with the desired tension. During the cooling, there is a difference in temperature between the ring and the glass, which cools faster than the ring. If the cooling of the glass is too rapid, the glass will not have adequate tension.

Weathering of the glass target, which occurs readily even in the relatively dry air-conditioned atmosphere of the factory, is a problem in two different ways. It shows up as smudge and also seems to be closely related to the problem of a sticking picture. The smudge problem has been fairly satisfactorily overcome by using a uniformly weathered surface produced by having a trace of sulfur in the box in which the mesh and ring heated for mounting the target. The sticking problem is not completedly solved but may require prevention of weathering altogether.

Dynodes

All of the dynodes in the image orthicon, except the first dynode, are of silver-magnesium alloy. The parts are processed (before they are assembled into the tube) to yield a surface film of magnesium oxide for high secondary emission. The processing consists of heating the parts by induction in a vacuum system. Oxidation is produced by water vapor obtained by allowing a liquid-air trap containing water to warm up. The magnesium is oxidized as it diffuses to the surface. The parts are cooled in air so that the magnesium remaining in the volume of the metal will also be oxidized. (The hot silver is permeable to oxygen but not to water vapor.) The first dynode of the image orthicon requires a particularly uniform and stable secondary-emission ratio because the beam scans across it. To obtain stability for the first dynode, it is necessary to use material having a lower secondary-emission ratio than that of the other dynodes.

Photocathodes

In general, the glass substrate for the semitransparent photocathodes should be very clean. The surface is first polished with cerium oxide to remove possible weathering and then is carefully washed.

While it is possible to evaporate the non-alkali metals for the photocathode before the main exhaust cycle, the sensitivities so obtained are not as high as when the complete photocathode is deposited during final exhaust.

The Ag-Bi-O-Cs photocathode is made by evaporating a silver-bismuth alloy to a given light-transmission value, oxidizing it to a second given transmission value, and then following with cesiation with the tube heated to about 165 C so that cesium vapor moves about. Generation of cesium is continued to a peak of light sensitivity.

The Sb-Cs photocathode is made by evaporation of antimony to a given transmission, followed by cesiation as described above. A small amount of oxygen is helpful and may be added after cesiation with the tube at room temperature. Another method is to evaporate manganese to a given transmission value and then to oxidize the coating with oxygen in a glow discharge to a second given transmission value before the deposition of the antimony. The manganese sublayer also helps obtain a satisfactory amount of lateral electrical conductivity in the Sb-Cs photocathode.

The multialkali photocathode is made by first evaporating antimony. The tube is then heated and is treated with the vapors of sodium, potassium, and cesium in a rather elaborate schedule. This photocathode is easily poisoned by small traces of oxygen and, apparently, by other gases. The bake portion of the exhaust cycle must be quite thorough to obtain maximum sensitivity.

REFERENCES

1. IRE Standard on Electron Tubes: Definitions of Terms, 1957 (57 IRE 7.52)
2. Fink, D. G., Television Engineering Handbook, pp. 5-3 to 5-81, McGraw-Hill, 1957
3. Zworykin, V. K. and G. A. Morton, Television, 2nd Ed., pp. 305 to 381, Wiley, 1954

4.* Weimer, P. K., "Television Camera Tubes: A Research Review," Princeton Technical Report 866 - To be published in Advances on Electronics and Electron Physics, Vol XIII, Academic Press

5.* Jones, R. C., "Quantum Efficiency of Detectors for Visible and Infrared Radiation," Advances in Electronics and Electron Physics, Vol. XI 1959, pp. 87 to 178, Academic Press

6. Rose, A. "Electron Optics of Cylindrical Electric and Magnetic Fields," Appendix A, Proc. IRE, 28, 38-39, 1940

7. Langmuir, D. B. "Theoretical Limitations of Cathode-Ray Tubes," Proc. IRE, 25, 977, 1937

8. Pierce, J. R. "Limiting Current Densities in Electron Beams," J. App. Phys., 10, 715, 1939

9. Thompson, B. J. and L. B. Headrick, "Space-Charge Limitations on the Focus of Electron Beams," Proc. IRE, 28, 318, 1940

10. Pierce, J. R., Theory and Design of Electron Beams. Van Nostrand, 1954

11. Schwartz, J. W. "Space-Charge Limitation on the Focus of Electron Beams," RCA Rev., 18, 3, 1957

12. Moss, H. "A Space Charge Problem," Wireless Engineer, 22, 316, 1945

13. Schade, O. H. "Image Gradation, Graininess and Sharpness in Television and Motion-Picture Systems," Journal of the SMPTE, 64, pp. 593-617, particularly pp. 602-605, 1955

14.* Rose, A., "Television Pickup Tubes and the Problem of Vision" Advances in Electronics Vol. I, pp. 131-166, Academic Press, 1948

*Items 4, 5 and 14 are the same publication. L. Marton is editor.

Image-Converter Tubes

R. G. Stoudenheimer

Lancaster

An image-converter tube is an electron tube which reproduces on its fluorescent screen an image of an irradiation pattern on its photosensitive surface.

An image-converter tube has three functional parts: (1) the photocathode which emits electrons in a density pattern corresponding to the incident illumination pattern, (2) an electron lens which accelerates the emitted electrons and forms an electron image of the incident illumination pattern, and (3) a phosphor screen placed in the plane of the electron image which converts the electron image into a visible picture.

The picture on the screen of an image-converter tube has no scanning lines, a characteristic which distinguishes its picture from that of television systems. The absence of scanning makes the external circuitry simple, and makes the potential resolution capability high. Usually, the limiting resolution (at least in the center of the picture) is established by the resolving capabilities of the phosphor screen. Hence, the resolving capabilities of the screen assume high importance in image-converter tubes. For this reason considerable effort has been given to the preparation of fine-grain, high-resolution phosphor screens.

External energy is introduced into the tube in accelerating the electrons before they hit the phosphor screen. Although a photon of light delivers only a few electron volts of energy to an electron to enable it to leave the photocathode, an external energy of many thousands of electron volts is delivered to the released electron from the external power supply. This augmenting of the electron's energy makes the image-converter tube an energy amplifier with a potential capability for image intensification, limited primarily by the efficiencies of the photoemissive and cathodoluminescent processes.

The electrons may be focused into an image by either electrostatic or magnetic lens action. In some cases where high resolution is not required, a simple accelerating field without focusing action can be used successfully. An electrostatic lens provides any desired magnification or demagnification, inverts the image, and is subject to all the aberrations encountered in light optics. Several systems of magnetic focusing may be used. The most common system employs superimposed uniform electric and magnetic fields. A modified version of this system employs magnetic and electric fields concentrated near the photocathode. Such a system provides a magnified, erect image with possibly a small amount of image rotation. The short magnetic lens operating in a region of high electron velocity, as used in cathode-ray tubes and the electron microscope, is ordinarily not used in image-converter tubes. Such a system provides an inverted image when the magnetic field is very short, and an inverted image which shows increasing rotation as the magnetic field is lengthened.

HISTORY

One of the first image-converter tubes was made by G. Holst[1] and others in the Philips Laboratories at Eindhoven and consisted of two parallel electrodes, a photocathode and a phosphor screen, spaced very close together. No focusing was used; the high field between cathode and screen prevented appreciable spreading of the electrons in traveling from a point on the photocathode to the screen.

Morton[2] introduced the image tube with an electrostatic lens and proposed a cathode with spherical curvature to minimize curvature of field and distortion. He also developed an image tube (type 1P25) which was used by the American Armed Forces in World War II. A similar but larger tube was developed by Dr. Walter Schaffernicht[3] (AEG Bildwander Tube) for use by the German Army.

After World War II, military emphasis in America was placed on developing high-voltage image tubes with ringseal construction (such as the 6032 and 6380) and unipotential image-converter tubes which could operate without a voltage divider and with a very small amount of power.

In addition to the use of image-converter tubes for military applications, tubes were developed for intensifying X-ray fluoroscope images, for light-shutter service, and with cascade arrangement for intensifying very faint images.

For X-ray intensification, tubes have been developed by Westinghouse, [4] Philips, [5] and the French Thompson Houston Company. Mullard[6] has developed a series of tubes with magnetic focus for use in light-shutter applications. Recently RCA[7] and DuMont[8] have developed image-converter tubes with a gating grid and deflecting plates for light-shutter service. An image-converter tube was developed by Bliss and Ruedy[9] for high-

speed printing in computers.

Morton, Ruedy, and others[10] at the RCA Laboratories analyzed the theoretical performance possible from cascaded-image tubes and built a few tubes which were evaluated primarily by military groups for performance with faint illumination. Similar tubes were made by Stoudenheimer, Moor, and Palmer[11] primarily for use in nuclear-track imaging. Russian scientists[12] have used cascaded-image tubes with magnetic focusing to photograph tracks of nuclear particles in the scintillation chamber. The Russian type of cascaded tube had sufficient gain for the photographic recording of single electrons released from the first photocathode.

L'Allemand[13] at the University of Paris has achieved excellent results by focusing the electron image directly on a photographic plate to obtain a factor of 100 increase in the effective speed of photographic film.

Sternglass[14] and others have developed an image-intensifier tube in which the primary photocurrent is multiplied many times by transmission secondary emission provided by a series of thin-film dynodes located in successive electron-image planes.

CONVERSION GAIN

The ability of the image-converter tube to intensify an image under certain conditions of operation is one of its most useful characteristics. The characteristic used to rate a tube for its ability to intensify an image is called the conversion gain. Generalized, the conversion gain is the ratio of the amount of exit flux to the amount of incident flux. Depending on the application, the incident and exit fluxes may vary in spectral distribution and are measured in various units. For direct visual observation, a user is interested in measuring the luminous-flux conversion gain. If the output screen is to be photographed, the user is interested in the total radiation from a blue screen and will therefore measure radiant-flux conversion gain.

The first practical image-converter tubes were used in the sniperscope to convert a near-infrared image into a visible image. The conversion gain measured for this application is the quotient of the exit luminous flux divided by the amount of incident near-infrared radiation. Gain is desirably, although not necessarily, a pure number without units. Consequently, to arrive at a gain figure which is dimensionless, it is necessary to specify the amount of near-infrared radiation (which is invisible) in terms of the amount of an equivalent luminous flux. The standard way of stating the amount of near-infrared flux is to give it in terms of the amount of luminous flux which would produce the same luminous output flux from the image tube as produced by the incident near-infrared flux.

The primary light source, both in actual use and in the test, is a tungsten lamp operating at a color temperature of 2870 K. For testing the tube, a stated amount of this flux is made to fall on the tube within a small central area of the photocathode. Measurements may then be made of the incident luminous flux F_i and the exit luminous flux F_e from the tube. If an infrared transmitting (IR) filter is placed between the source and the image tube, the incident luminous flux is reduced essentially to zero but the output screen flux is now F_eT, where T is the effective transmission of the IR filter. For the observer, viewing the screen only, the effect of introducing the filter is the same as if the incident luminous flux were reduced to the value F_iT without a change in the spectral distribution of the incident flux. Consequently, the incident infrared radiation can be stated to be equivalent to F_iT lumens from a tungsten lamp operating at 2870 K. The conversion gain (called conversion index CI for this particular test) is:

$$CI = \frac{\text{measured output flux}}{F_iT}$$

The filter transmission T is not measured for each individual tube tested but is calculated for a tube having the true S-1 response* by using the equation:

$$T = \frac{\int R_\lambda \, J_\lambda \, T_\lambda \, d\lambda}{\int R_\lambda \, J_\lambda \, d\lambda}$$

where R_λ = the value of the JEDEC S-1 relative spectral response at wavelength λ

J_λ = the value of the relative spectral emission of the 2870 K light source at wavelength λ

T_λ = the filter transmission at wavelength λ

The conversion gain of image-converter tubes may be made very high by cascading them. In cascaded-image tubes, the cathode of the second stage is separated from the phosphor screen of the first stage by a very thin transparent membrane less than 0.001-inch thick. The total conversion gain of a two-stage image tube may be 1000 or more.

The theoretical conversion gain can readily be calculated to compare with measured values. Calculation of the theoretical infrared conversion index requires knowledge of the cathode infrared sensitivity S_{ir} (in terms of the amperes per lumen of flux incident on the infrared filter), the filter transmission factor T, and the luminous conversion efficiency of the screen E. In practice for a Corning 2540 filter and P-20 screen, where typically: $S_{ir} = 4 \times 10^{-6}$ ampere per lumen

$T = 0.116$

$E = 50$ lumens per watt

$V =$ applied voltage of 16,000 volts

$$CI = \frac{S_{ir}VE}{T} = \frac{4\times10^{-6}\times1.6\times10^{4}\times50}{0.116}$$

$$CI = 27.6$$

The luminous flux conversion gain G for a two-stage cascaded image tube is

$$G = s_1V_1E_1K\sigma_2V_2E_2$$

*The JEDEC spectral response designation describes the overall characteristic of the tube including modifications by the transmission properties of the envelope; the designation in no way characterizes properties of the photosurface alone.

Table I

Definitions and Typical Values of Gain Symbols

Symbol	Definition	Typical Value
s_1	effective luminous sensitivity of first photocathode	140×10^{-6} amp/lumen
σ_2	peak radiant sensitivity of second photocathode	0.05 amp/watt
V_1, V_2	voltage across first and second stages, respectively	14,000 volts
E_1	efficiency of first screen	0.1 radiated watts/watt
E_2	luminous efficiency of second screen	50 lumens/watt
K	factor for converting peak radiant sensitivity of second cathode to sensitivity for P-11 radiation	0.88 for S-20 response

Generally the first photocathode has effectively S-20 spectral response, the first screen has P-11 spectral energy radiation, the second cathode has effectively S-20 response, and the final screen has P-20 spectral energy radiation. If the symbol definitions and typical values given in Table I are inserted in the gain equation, then,

$$G = (1.4\times10^{-4})\times(1.4\times10^{4})\times(0.1)\times(0.88)\times(0.05)\times(1.4\times10^{4})\times(50)$$

$$G = 6000$$

In the calculation of the theoretical gain of cascaded-image tubes, the integration of a product of a spectral-response curve and of a phosphor spectral-emission characteristic is usually encountered.

If $W_\lambda \, d\lambda$ is the radiated flux from the first screen between wavelengths λ and $\lambda + d\lambda$, and $\sigma_{2\lambda}$ is the radiant sensitivity at wavelength λ of the second cathode, then the effective radiant sensitivity $\bar{\sigma}_2$ of the second cathode to the phosphor radiation is

$$\bar{\sigma}_2 = \frac{\int \sigma_{2\lambda} W_\lambda \, d\lambda}{\int W_\lambda \, d\lambda}$$

but $\sigma_{2\lambda} = \sigma_2 R_{2\lambda}$

where $R_{2\lambda}$ is the relative spectral response normalized to unity at the wavelength of maximum response. Then

$$\frac{\bar{\sigma}_2}{\sigma_2} = \frac{\int R_{2\lambda} W_\lambda \, d\lambda}{\int W_\lambda \, d\lambda} = K$$

The units used for W_λ are entirely arbitrary since the same units appear in both numerator and denominator. The value of $\frac{\bar{\sigma}_2}{\sigma_2}$ has been calculated for a number of phosphor-cathode combinations; the results are shown in Table II.

Table II

Sensitivity of Various Photocathodes
In Glass Enclosures to Radiation
from Various Phosphors

Phosphor No.	Photocathode Response No.	$K = \bar{\sigma}_2/\sigma_2$
P-11	S-11	0.91
P-20	S-11	0.42
P-20	S-20	0.58
P-11	S-20	0.88

The screen efficiency is not a constant for all values of operating voltage. At low voltages the aluminum film, which is applied to the phosphor, stops the electrons; even without the aluminum film, the phosphor permits such shallow electron penetration at low voltages that the light scatters and is absorbed by the screen layer between the excited layer and the exit face. At very high operating voltages, electrons penetrate completely

through the screen without delivering all of their energy to the screen. A typical screen characteristic for a settled P-20 phosphor is shown in Fig. 1 where the maximum screen efficiency (slope of line from point on curve to origin) is obtained at about 16 kilovolts. High-efficiency screens can be made with particle sizes as small as one micron. If the particle size is appreciably smaller than one micron, the screens become somewhat too opaque when their thickness is great enough to absorb the energy of 20-kilovolt electrons and, as a result, screen efficiency is low.

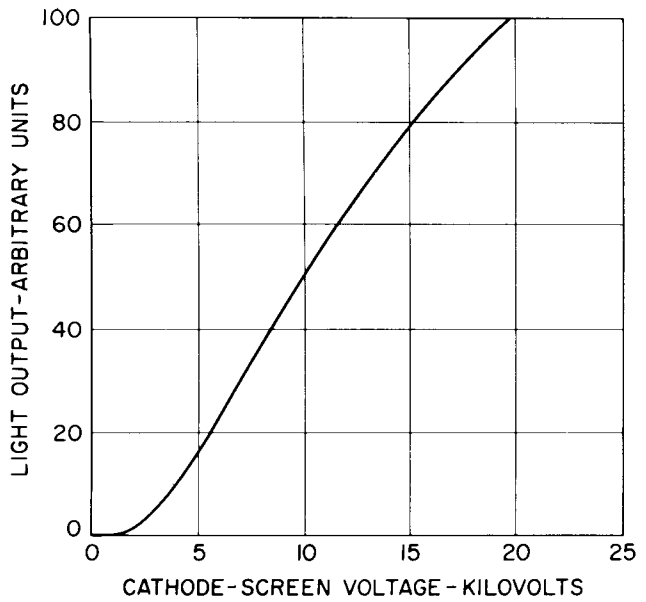


Figure 1. Light Output-vs-Voltage Characteristics of P-20 Screen

The first attempts to make infrared-sensitive image-converter tubes with sulfide phosphor screens resulted in photocathodes with slumping sensitivity to near-infrared radiation. This trouble arose because sulfide phosphors are easily damaged by baking at high temperatures (above 300 C) and, therefore, the first tubes were not baked above this temperature. However, the screens, especially the silicates used in these screens, are not thoroughly degassed at temperatures below 450 C. As a result, if the tubes are baked at temperatures below 450 C, the silver-oxygen-cesium (Ag-O-Cs) photocathode apparently absorbs gases released from the screen during cathode processing and becomes contaminated with materials that have a great affinity for adsorbing excess cesium. Although the photocathodes of the early tubes had good sensitivity immediately after their preparation, they lost sensitivity on standing. This loss in sensitivity would start in the worst tubes after a few days, and in the better tubes after two or three weeks. The normal high sensitivity could be restored by baking the tubes at a temperature which drives the excess cesium off the surface of the photocathode. The slumping could result from cesium which diffuses to the surface of the cathode from its interior or from cesium transferred as a vapor from other parts of the tube. Slumping sensitivity was only eliminated by the use of a long degassing bake at a temperature of 450 to 460 C.

If a lens forms an image on the photocathode of an image-converter tube the ratio of screen brightness to scene brightness is the brightness gain of the system.

The illumination E on the tube is

$$E = \frac{B\,T}{4(f/d)^2(1 + M^2)}$$

and the screen brightness B' is

$$B' = \frac{E\,G}{m^2}$$

The brightness gain $\frac{B'}{B}$ is

$$\frac{B'}{B} = \frac{T\,G}{4(f/d)^2(1 + M)^2\,m^2}$$

where B = scene brightness
T = transmission of lens
f/d = f number of lens
M = magnification of image formed by lens
G = conversion gain of tube
m = tube magnification

SCREEN BACKGROUND

If an image-converter tube is placed in complete darkness, a residual faint glow will be seen on the output phosphor. This residual glow is called screen background. Screen background lowers contrast at low levels of incident illumination and limits the application to illumination levels where the screen brightness is well above that of the background. High screen background may be due to thermionic emission from the photocathode or to field emission. Field emission, even if it does not result in direct electron bombardment of the screen, may produce fluorescence in a bombarded region of the glass bulb; the light from the bombarded glass is fed back to the cathode where it generates photoelectrons that increase the screen background.

Thermionic Emission

In a well-designed and well-constructed image-converter tube, background is due only to thermionic emission from the photocathode. In a photodevice having S-1 response, it is difficult to reduce the thermionic emission much below 10^{-12} ampere per square centimeter at room temperature. With a device having a cathode sensitivity of 30 microamperes per lumen, the thermionic emission of the cathode is equal to the photocurrent produced by an illumination of 3.3×10^{-5} millilumens per square centimeter (3.1×10^{-5} foot-candles).

In the absence of screen background,

$$\text{screen brightness} = \frac{\text{cathode illumination} \times \text{conversion gain.}}{(\text{tube magnification})^2}$$

Consequently, for a given conversion gain and tube mag-

nification there is a value for the cathode illumination which would produce a screen brightness equal to the background brightness.

$$\frac{\text{Screen background brightness} \times (\text{tube magnification})^2}{\text{conversion gain}} =$$

Screen background equivalent cathode illumination.

This expression is constant where thermionic-emission density is constant. In a given tube type, magnification is constant and the test for screen background is established as a test for the quotient of conversion gain divided by screen background.

$$\frac{\text{conversion gain}}{\text{screen background}} = \frac{(\text{magnification})^2}{\text{equivalent cathode illumination of screen background}}$$

This quotient may be evaluated for type 6032, which has a tube magnification of 1/2. Assuming a thermionic emission of 10^{-12} ampere per square centimeter and a cathode sensitivity of 30 microamperes per lumen, the equivalent cathode illumination of screen background can be calculated to be 0.033 microlumen per square centimeter.

$$\frac{\text{conversion gain}}{\text{screen background}} = \frac{1/4}{0.033} = 7.5 \text{ per microlambert}$$

A practical minimum limit of 2.0 per microlambert was established for production testing. In tubes where the magnification is 0.75, the minimum test limit in common use is 3.0 per microlambert.

When type 6032 was first placed in production, screen background brightness was high primarily because of high thermionic emission from the Ag-O-Cs photocathode. Since slump in infrared sensitivity was observed in some tubes at the same time, there was a good opportunity to compare the relative change in thermionic emission with the infrared sensitivity. In these tubes, the background due to thermionic emission was observed to decrease by almost two orders of magnitude while the infrared sensitivity* dropped to about one-third of the initial values (see Fig. 2). The threshold shifted toward shorter wavelengths as the tubes slumped. These changes which accompanied slump were much the same as those found by Prescott and Kelly[15] as the adsorbed cesium on the photocathode increased.

Evaporation of silver on a photocathode after exhaust is effective in controlling infrared sensitivity and thermionic emission. A very small amount of evaporated silver is required for the cathode sensitivity to reach a peak and start decreasing unless the cathode is heated to a temperature of about 150 to 175 C after the silver evaporation. Heating probably allows cesium to diffuse from the interior to the surface of the photocathode until equilibrium is established with the interior. The use of the term "resilvering" here implies a subsequent heating after silver evaporation to restore equilibrium with the interior. As the silver on the surface is slowly increased by resilvering, both the infrared sensitivity and the thermionic emission increase as shown in Fig. 3, then decrease with further resilvering.

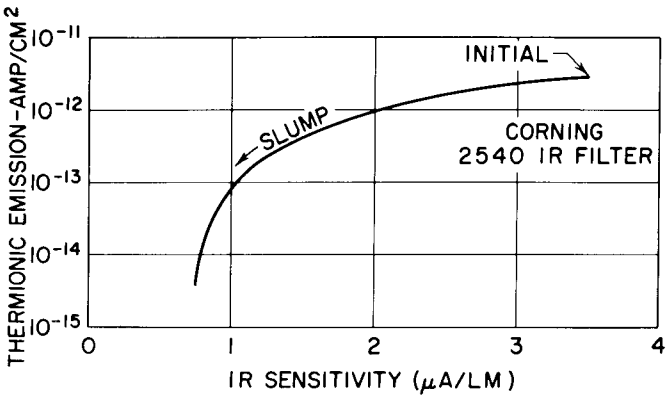


Figure 2. Relation between Infrared Sensitivity and Thermionic Emission as Cathode Sensitivity Slumps

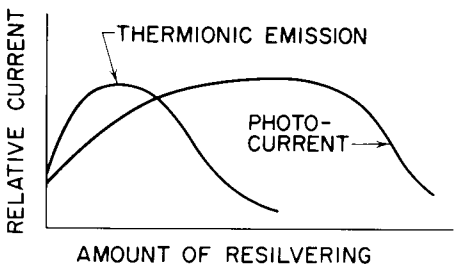


Figure 3. Effect of Resilvering on Infrared Sensitivity and Thermionic Emission

By proper control of the amount of resilvering, high cathode sensitivity and low thermionic emission are possible. If the resilvering is stopped earlier, the cathode sensitivity may be equally good but thermionic emission is quite high. With proper resilvering of a cathode in an image-converter tube with S-1 response, a thermionic emission of 10^{-12} ampere per square centimeter may be obtained. It is also possible to control the resilvering to obtain a thermionic emission as high as 10^{-10} ampere per square centimeter.

Thermionic emission is a fundamental cause of screen background which cannot be reduced by good design. If a low screen background is desired, a cathode with low thermionic emission should be used. The cathodes listed in Table III are in the order of decreasing thermionic emission.

Field Emission

Field emission is also a frequent cause of high screen background. Where field emission has been serious in the past, one of the following conditions was found responsible.

*Infrared sensitivity is defined as the quotient of the output current divided by the luminous flux incident on the infrared filter, where all radiation passing through the filter is incident on the photocathode. Light source is a tungstem lamp operating at a color temperature of 2870 K.

TABLE III

Sensitivity and Thermionic Emission for Various Cathodes

Composition of Photosurface	Sensitivity to 2870 K light source — microamperes/lumen	Thermionic Emission at 20 C — amperes/sq. centimeter
Silver-Oxygen-Cesium	30	10^{-12}
Silver-Oxygen-Bismuth-Cesium	60	7×10^{-14}
Antimony-Cesium-Oxygen	60	3×10^{-15}
Antimony-Sodium-Potassium-Cesium	150	2×10^{-16}

Negative Electrodes with Sharp Edges. The example used here is the 1C6-2, an early developmental version of type 6929 developed for the U. S. Army Engineer Research and Development Laboratories. The tube is shown in Fig. 4.

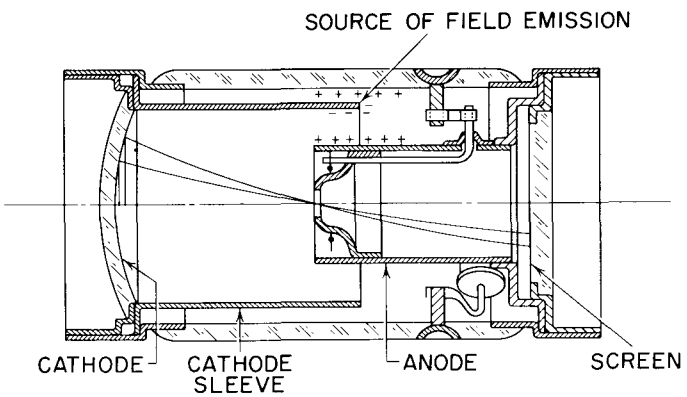


Figure 4. IC6-2 Developmental Version of Type 6929 Image-Converter Tube

At the edge of cathode sleeve, high fields were established which resulted in field emission at a voltage of about 6 kilovolts. The field-emission electrons did not usually bombard the screen directly, but caused high screen background indirectly by bombarding the glass wall which fluoresced and resulted in light feedback to the cathode. When a tube is similar in construction to that shown in Fig. 4, it is desirable to keep the radius of curvature of the edge of the cathode cylinder as large as possible. Various types of donut-shaped rings (sometimes called corona shields) are frequently attached to this type of electrode. Usually this type of construction, even with corona shields, is not as successful for high voltage operation as other designs.

Glass-to-Metal Seals Subject to Electron Bombardment. If glass is subject to electron bombardment, it charges to a high potential because of the secondary electrons emitted. Frequently, as the glass charges, the electron bombardment increases and the charged area spreads towards a negative electrode to which the glass is sealed. When the highly charged area reaches a negative electrode to which it is sealed, tremendous fields are established between the edge of the glass and the metal and cause very high field-emission currents which exceed the current available from the power supply. As a result, the disappearance of voltage across the tube causes the discharge to appear as a flash.

The early developmental versions of type 6032 (see Fig. 5) were subject to this type of breakdown. Bombardment of the glass by thermal electrons from the grid were sufficient to initiate charging of the glass surrounding the anode. The charged area would soon creep to the grid and cause a burst of field emission from the grid at the boundary between glass and metal. To eliminate this type of breakdown, it is necessary to shield the glass from electron bombardment. If glass charges randomly in the region close to a gap in the electrodes, the resulting charges deflect the beam and cause the image to shift position on the viewing screen. Because of the magnification in the viewing lens, a very small shift in image position is quite annoying. Field emission at glass-to-metal seals may sometimes be effectively reduced by the addition of a thin film of glass over the negative electrode close to the seal area. Electrical conduction through the thin glass prevents charging if bombarding currents are not too large and a large excess of cesium is avoided.

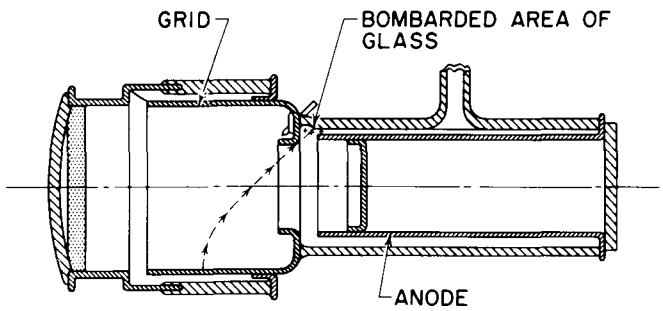


Figure 5. Developmental Version of Type 6032 Tube Showing Glass-Charge Effects Adjacent to Sealed-to-Glass Negative Electrode

Excess Cesium. Excess cesium used for processing the cathode can lower the work function of negatively charged surfaces and thus increase field emission. The highest electric-field gradients causing field emission are found at the electrode closest to the anode. Consequently, excess cesium should be avoided in this region. Considerable trouble with field emission has been encountered in a light-shutter type of image tube when the cesium pellets were located on the anode. However, when the cesium source was moved to a position very close to the cathode, the operating voltage could be doubled without trouble from field emission. Excess cesium increases field emission both at sharp edges and at glass-to-metal seals. Residual gas may also increase screen background. However, residual gas is more troublesome because it results in screen excitation proportional to the cathode photocurrent and causes a general loss of contrast over the picture.

ELECTRON LENS SYSTEMS

The quality of the image appearing on the output phosphor screen is of fundamental importance in image-converter tubes. The quality of this image is determined

by its resolution on the axis, its resolution in different zones off the axis, and the magnitude of the distortion in the image.

Resolution may be limited by the thickness or grain of the phosphor screen, or it may be limited by the aberrations of the electron lens system. Distortion arises because the electron-lens system does not image the electron source in its true shape. The distortion is usually observed as a difference either in linear magnification or in image rotation between the central and outer zones of the image.

The electron-lens system used in image-converter tubes differs from that used in most other electron-imaging devices in that the lens system must be a wide-angle system (where the image is inverted), or it must cover a large image generally whether the image is inverted or erect.

Uniform Electric and Magnetic Fields

The most common magnetic-lens system for image tubes employs uniform electric and magnetic fields. Basically, the system is similar to that shown in Fig. 6.

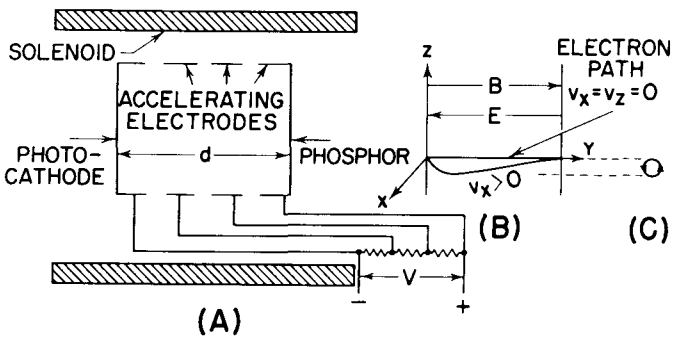


Figure 6. Magnetically Focused Image Converter: (A) schematic arrangement of electrodes, (B) magnetic and electric field and paths of electrons with and without a vertical velocity component, (C) electron paths as seen from the screen

The accelerating electrodes are generally equally spaced along the axis. The voltage increments are essentially the same between all electrodes, but should be adjusted to provide a uniform electric field from cathode to screen. A solenoid is placed around the tube to provide the uniform magnetic field. The solenoid should be longer than the tube to avoid the weaker, flared, magnetic field at the ends. End effects in the coil may be minimized by adding more turns at the end. A permanent cylindrical magnet may also be used to provide a uniform magnetic field.

If the voltage between the cathode and screen is V, and the distance between them is d, the transit time t of an electron emitted with zero axial velocity is:

$$t = d\sqrt{\frac{2m}{eV}}$$

where m = the mass of the electron

and e = the charge of the electron

If an electron is emitted with a small component of velocity perpendicular to the magnetic field, this electron will travel in a spiral path towards the screen and require a time T to complete a revolution around a magnetic line of force, thus

$$T = \frac{2\pi m}{Be}$$

where B is the magnetic field strength

If t = T, the spiral turn is complete when the electrons arrive at the screen, all electrons leaving a point on the cathode are focused into a point image on the screen, and the following relation for focus exists:

$$d\sqrt{\frac{2m}{eV}} = \frac{2\pi m}{Be}$$

$$d = \sqrt{\frac{e}{2m}}\ \frac{2\pi m}{e}\ \frac{\sqrt{V}}{B}$$

when d is in centimeter, V is in volts, and B is in Gauss

$$\text{then } d = 10.58\ \frac{\sqrt{V}}{B}$$

The image is erect and has the same size as the electron source.

Short Magnetic Lens

The long magnetic lens represents one extreme case of the magnetic lens which is relatively easy to analyze. The other extreme case is the hypothetical very short axial magnetic field which has an axial component H_z over a very short distance but has no radial component. If this magnetic field is placed in a region where the electric potential is constant and equal to ϕ_o, it will behave as a lens where the focal length f is given by:

$$\frac{1}{f} = \frac{e}{8m\phi_o}\int_A^B (H_z)^2\, dz$$

Object and image distances are related to focal length in the same manner as in light optics. If a real image is formed, it is inverted. Magnification depends on the ratio of the image distance to the object distance. Although a very short magnetic lens can be used in image tubes, off-axis aberrations would be very severe unless the focus-coil diameter were quite large. This type of lens has found very little application in image tubes, but is widely used in cathode-ray tubes and in the electron microscope where the object field is small. One special type of short magnetic lens, having more resemblance to the long magnetic lens than to the typical short lens, has been frequently used in image tubes. It was found that image defects were minimized, if a short-lens coil was made in such a way that the flux lines close to the cathode approximately coincided with the paths taken by the electrons in the absence of the magnetic field.

A tube developed by Iams, Morton and Zworykin[16] is shown in Fig. 7. The lens in this tube produces a real image which is rotated about 30 degrees with respect to the object. The electrons do not cross the axis of the system as they do in the very short magnetic lens placed in a region of constant high electric potential. The electric field inside the anode forms a negative electrostatic lens which causes the electrons to diverge away from the axis. This divergence (along with the diverging magnetic field in the same region) causes the image to be magnified. However, because all electrons start their path parallel to the axis at the cathode and diverge only slightly inside the anode, the variation in convergence is small for electrons starting from different zones of the cathode. Consequently, image curvature is small and image resolution is good to the edge of the picture.

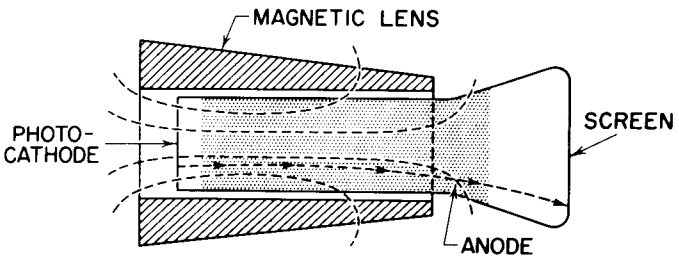


Figure 7. Image Tube with Special Short Magnetic Lens

The Electrostatic Lens

The electrostatically focused image-converter tube is probably the most widely used. It is used in military equipment because no focus coil or magnet is required and equipment may be made compact and light. Power requirements for the electrostatically focused tube may also be very small, especially if the tube uses a single voltage (monovoltage tube) and can, therefore, be used with a power supply having no voltage divider.

The fundamental electron lens system which provides a crude axial image is as simple as the double-convex lens in light optics. However, a lens system free of aberrations is more complex and difficult to attain. Basically, two cylindrical electrodes having a common axis and a voltage between them (as shown in Fig. 8) form an electron lens. Within the limits which permit real images, there exists, for any electron source along the axis inside the negative electrode, a corresponding real image point along the axis inside the positive electrode. However, for off-axis points, such a lens is found to have a large amount of astigmatism and a highly curved image field. Morton and Ramberg[2] analyzed the relations existing between lens diameter, axial object distance, and axial image distance; these relations are shown in Fig. 9. Morton found that astigmatism and curvature of image field could be greatly reduced by making the cathode curved with the center of curvature near the desired axis-crossover point. Morton also showed that, in all electrostatically focused image tubes, the tube magnification is approximately equal to one-half the ratio of image distance to object distance.

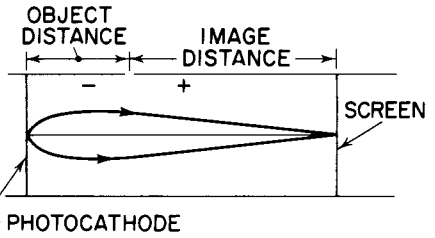


Figure 8. Simple Electrostatic Lens System

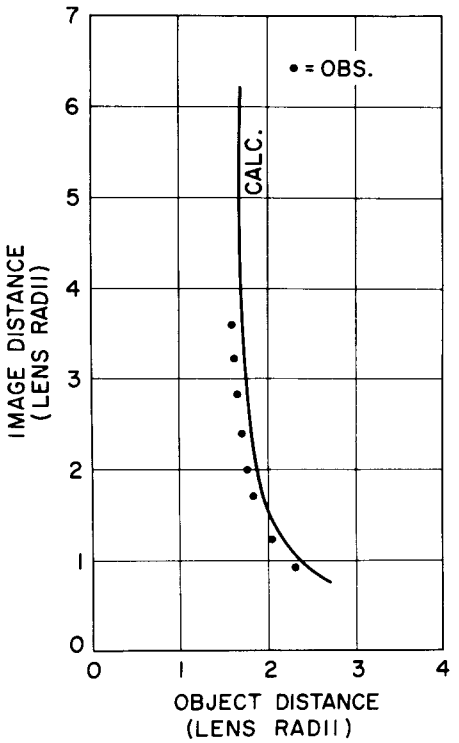


Figure 9. Relation between Object Distance and Image Distance in a Simple Electrostatically Focused Image-Converter Tube

In electron optics, as in light optics, image distance and object distance are related by the expression:

$$\frac{n}{s} + \frac{n'}{s} = \frac{n'}{f'} = \frac{n}{f}$$

where n = index of refraction in object space (corresponding to $\sqrt{\text{potential}}$)
n' = index of refraction in image space
s = object distance to first principal plane
s' = image distance to second principal plane
f = focal length in object space
f' = focal length in image space

Unfortunately, any change in the position of the photocathode changes the field distribution so that the focal length of the electron lens is changed. If an attempt is made to change the object distance it is almost impos-

sible to estimate how much the focal length of the lens is changed at the same time. Consequently, determining image position and tube magnification before building a tube requires measurements of the axial potential distribution and calculation of paraxial electron trajectories. The electron trajectories can be calculated by solving the paraxial-ray equation:

$$\frac{d^2r}{dz^2} + \frac{\phi_o' dr}{2\phi_o dz} + \frac{r\phi_o''}{4\phi_o} = 0$$

Methods of solving the paraxial-ray equation are given by Maloff and Epstein,[17] Spangenberg and Field,[18] and Gans.[19,20] Gans' method is simple and reasonably accurate. Calculation of two electron paths is sufficient to locate the paraxial image and determine paraxial magnification. The paraxial image is located by starting an electron at the cathode on the tube axis in a direction oblique to the axis and tracing the path until the electron returns to the axis. An initial electron energy of one electron volt and a velocity vector making an angle of 45 degrees with the axis are convenient starting conditions. A second electron is started a short distance (object height) off the axis in a direction parallel to the axis. This electron crosses the axis and intercepts the paraxial image plane on the opposite side of the axis at a distance from the axis equal to the image height. The ratio of image height to object height is the tube magnification.

Fig. 10A shows a monovoltage image tube designed to provide a given magnification with a given distance from cathode to screen without a special electrode for adjusting the electrical focus. Various models were constructed in an electrolytic tank where the potential distribution along the axis was measured. A typical potential plot is shown in Fig. 10B. Gans' method was used to trace the required electron paths (Fig. 10C). By proper adjustment of the convergence of the electron lens and by placement of the lens at the proper distance from the cathode the electrons could be imaged on the screen with the desired magnification.

Lens Aberrations

In both light optics and electron optics, two aberrations appear in the paraxial image and four additional aberrations appear in images removed from the axis. These are (1) paraxial aberrations, including chromatic aberration and spherical aberration, and (2) off-axis aberrations including coma, astigmatism, curvature of field, and distortion. In the following discussion, major emphasis is given to the aberrations encountered in electrostatically focused image tubes.

Chromatic Aberration. This aberration is due to a variation in index of refraction for light of different wavelength or for electrons of different speed. The aberration is usually very small in electron optical systems where the spread in initial electron velocity is small compared with maximum electron velocity; it is small where a photocathode is the electron source, but it may be a serious limitation in systems using secondary emitters as an electron source. Henneberg and Recknagel[21] derived formulas for the chromatic aberration in electron lenses and obtained the following formula for image tubes with either an electrostatic lens or a uniform electric and magnetic field.

Diameter Δ (in millimeters) of circle of diffusion on screen to a first approximation is

$$\Delta = 2m\frac{V}{E_k}$$

where m = tube magnification
V = energy of emitted electron in electron volts
E_k = electric field strength at the cathode in volts per millimeter

thus, for a type 6032 image tube,
if m = 0.5
V = 0.5 electron volts
E_k = 50 volts per millimeter

Then $\Delta = 2 \times 0.5 \left(\frac{0.5}{50}\right) = 0.01$ millimeter

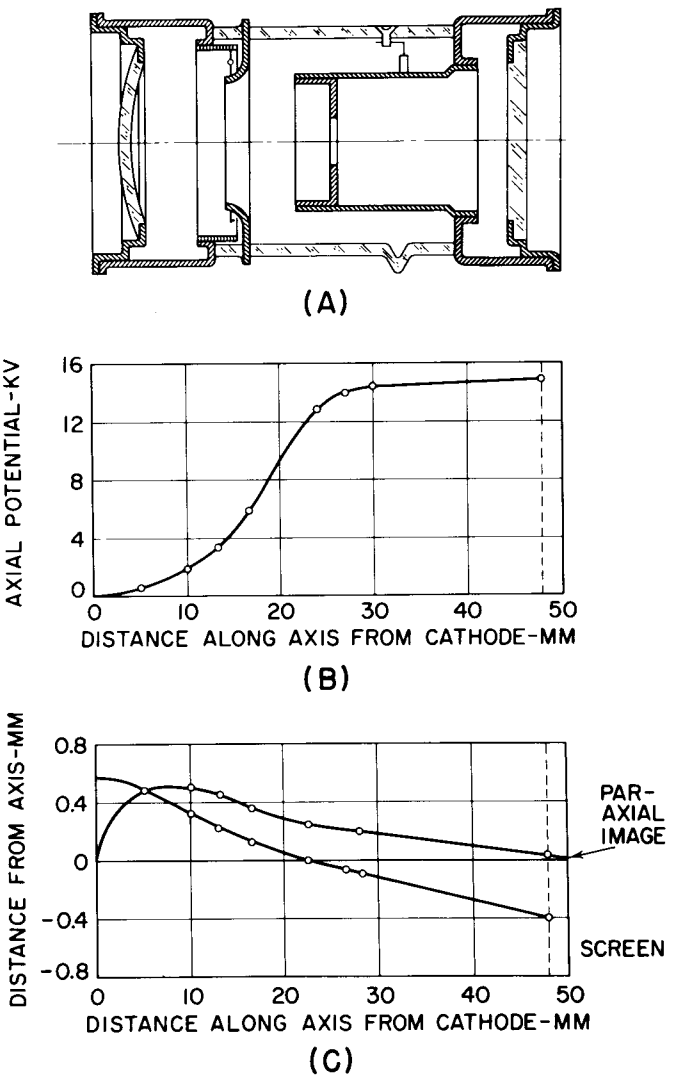


Figure 10. Potential Plot and Electron Paths of Monovoltage Image Tube: (A) developmental type C73413A image tube, (B) axial potential distribution, (C) calculated electron paths, using Gans' method

With a P-20 screen, electron images smaller than 0.02 millimeter are not resolved and, therefore, chromatic aberration will not limit resolution for this type of screen. With an evaporated transparent screen of willemite, 150 lines per millimeter have been resolved. The latter resolution is somewhat better than theory predicts, but if a large number of electrons are emitted with an energy of less than 0.5 ev, the distribution of electrons in the disk of diffusion might be such that two images spaced closer than 0.01 millimeter might still be resolved.

Spherical Aberration. This aberration arises because of zonal variations of the focal length of a lens. Although the electrostatic type of electron lens in an image tube has serious spherical aberration, the small diameter of a bundle of electrons from a point source results in a small effective aperture of the lens and very little spherical aberration in the image. Spherical aberration could become serious at low electrode voltages where the spread of the electron bundle is large. Spherical aberration is related to other aberrations and the methods of controlling its magnitude are therefore important. At the present time, it is impossible to eliminate spherical aberration in either a single-component spherical optical lens or in any electron optical-lens system. In optical lenses with spherical surfaces, spherical aberration is controlled by the shape (bending) of the lens. All simple types of convergent lenses have positive spherical aberration, that is, the focal length for outer zones is shorter than the focal length of the central zone.

Coma. Coma, like spherical aberration, depends on the effective aperture of a lens. Because an electron bundle in an image tube is so small in diameter, coma is not a serious aberration in these tubes.

Curvature of Field and Astigmatism. These two aberrations are the most serious aberrations now encountered in those image tubes which invert the electron image. They are closely related and can best be discussed if grouped together. In the case of all simple lenses used in light optics, it was realized very early that the image field is not flat but curved towards the lens. In the absence of astigmatism, the radius of curvature is given by a theorem[22] commonly ascribed to Petzval, but actually known and proved before this time. This theorem states that, in the absence of astigmatism, the radius of curvature in the central part of the field is given by ρ, where $1/\rho = \sum [(n - n_0)/rnn_0]$, where r is the radius of curvature of a refracting surface separating media of indices n_0 and n, respectively. The manner in which curvature of field is introduced is illustrated in the simple lens shown in Fig. 11A where an aperture is placed at the center of curvature of the refracting surface to limit the imaging pencil of rays from an object at infinity to a narrow bundle along the radius and thus eliminate astigmatism. The focal length f of the refracting surface is given by the equation

$$(n - n_0)(1/r) = \frac{n}{f}$$

$$\frac{\Delta n}{nr} = \frac{1}{f}$$

$$f = \frac{nr}{\Delta n}$$

$$R_n = f - r = \frac{nr - r\Delta n}{\Delta n} = \frac{r(n - \Delta n)}{\Delta n} = \frac{rn_0}{\Delta n}$$

$$\frac{1}{R_n} = \frac{\Delta n}{n_0 r}$$

where $\Delta n = n - n_0$

R_n = radius of curvature of image in medium of index n

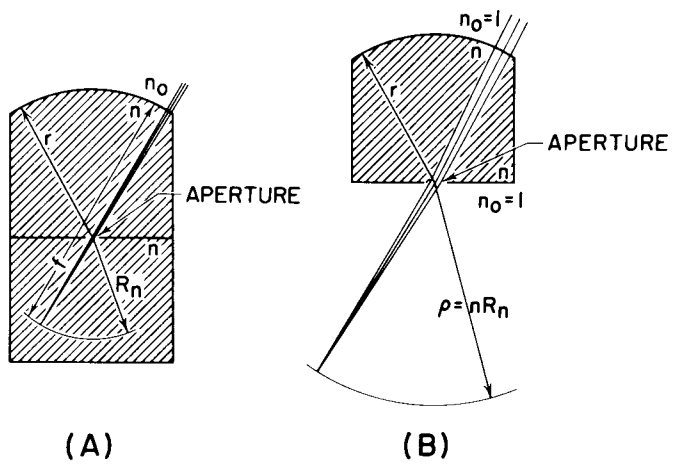


Figure 11. Curvature of Image Field, (A) Image in second medium, (B) A second (plane) boundary at the aperture permits the image to be formed in the first medium of index 1

If the medium of index n extends only to the plane of the aperture and the index of refraction of the medium after the aperture is unity, as shown in Fig. 11B, the image is in the latter medium and it is moved back so that the optical path beyond the aperture is the same as shown in Fig. 11A, where the medium after the aperture has an index of n.

Then the radius of curvature of image = $nR_n = rn_0n/\Delta n$ and $1/\rho = \Delta n/n_0nr$, which is the value predicted by the Petzval theorem where the image is assumed to be in air. The simple illustration is not a proof of the Petzval theorem. The latter applies universally to all lens systems except perhaps the anastigmatic lenses where the Petzval radius of curvature becomes very large. In these lenses, high-order deviations generally result in a Petzval surface with a reversed curvature near the edge of the useful field.

The contribution $\Delta(1/\rho)$ to Petzval curvature of a refracting surface in a system of many refracting surfaces is given by the equation

$$\Delta\left(\frac{1}{\rho}\right) = \left(\frac{1}{r}\right)\frac{n}{n_0 n}$$

*A single refracting surface freed of astigmatism by proper location of the aperture has a curved image field with the center of curvature at the aperture. The image lies in the Petzval surface.

The power* of the same refracting surface is given by the equation

$$P = \frac{1}{r} \Delta n$$

Because n_o, n, and Δn are positive in electrostatically focused image tubes, any values of r and n which increase the power of a lens must also increase the Petzval curvature.

One of the primary considerations in the design of camera lenses is the attainment of a very small Petzval sum. The Petzval sum is reduced in light optics by several practical methods. First, in a single meniscus-shaped lens, r_1 may be made equal to r_2. The negative power of the concave surface does not completely nullify the positive power of the convex surface if the lens is made sufficiently thick. Second, positive and negative lenses may be separated. When positive and negative lenses are placed closely together, their power is very small but when they are separated the required distance to eliminate the various aberrations, the power is increased considerably. Third, field-flattener lenses are often used in lens systems to control curvature of image. Where a lens has been corrected for all aberrations except curvature of field, a strong negative lens may be placed very close to the image plane, where it reduces the Petzval sum but does not contribute appreciably to the power of the lens system nor to any of the other aberrations.

The radius of curvature of the lens surface and the aperture location were chosen in Fig. 11 to eliminate astigmatism. Astigmatism is absent if a small bundle of incident rays is symmetrical about a radius generating the refracting surface. However, if a bundle of rays is oblique to the radius, the lens is stronger for those rays in the plane containing the axis and the object point, and, consequently, astigmatism is introduced. In this case, if the lens is positive, astigmatism is positive, while, if the lens is negative, astigmatism is negative. If a lens is afflicted with astigmatism, the Petzval theorem does not give the radius of curvature of the image but instead gives the radius of curvature of the abstract surface called the Petzval Surface.

If a lens has positive astigmatism, as shown in Fig. 12, the image field has more curvature than the Petzval surface P and is not well defined in a single surface. A point off the axis in object space is imaged as a small line tangent to a circle about the axis in the tangential image surface T, and as a small line along a radius towards the axis in the sagittal image surface S. Between the tangential and sagittal fields, the image is elliptical and provides the best resolution where the image becomes circular (the circle of least confusion). The distance of the tangential image surface from the Petzval surface is always three times the distance of the sagittal surface from the Petzval surface.

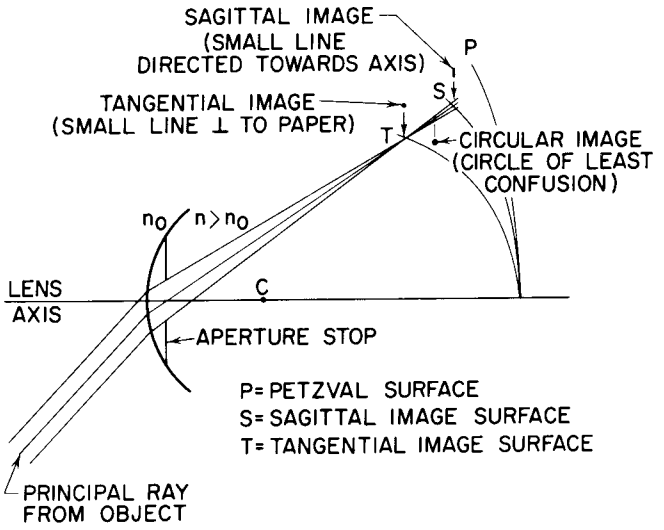


Figure 12. Convergent Refracting Surface Afflicted with Positive Astigmatism (An object off the axis but in the plane of the paper is imaged by rays in the plane of the paper at a shorter distance from the lens than by rays in the plane containing the principal ray but perpendicular to the plane of the paper.)

In a light-optical system having a fixed Petzval sum (Petzval curvature), astigmatism and, consequently, curvature of image field may be varied by changing the spacing between lenses or by changing the location of the aperture stop. *

The simplest example of a properly located stop to flatten the image field is the meniscus camera lens with a stop on the concave side of the lens as shown in Fig. 13. The center of curvature of the convex surface is beyond but close to the stop, while the center of curvature of the concave surface is beyond both the center of curvature of the convex surface and the stop. Examination shows that the principal ray (central ray of imaging pencil) crosses the convex surface at nearly normal incidence but crosses the concave surface at a large angle of incidence. Consequently, the concave surface introduces large negative astigmatism while the convex surface introduces little or no positive astigmatism. The residual negative astigmatism forms an image with curvature in both the tangential and sagittal images which is less than the Petzval curvature. The residual negative astigmatism increases as the bending of the lens increases provided the stop is located so that the principal ray, in passing through the convex surface, approximately coincides with the radius of curvature of the convex surface.

In electrostatically-focused image tubes now being

*The power of any lens is the quotient of the index of refraction of image space divided by the focal length in image space. The power P of a system of two lenses of power P_1 and P_2, respectively, is:

$$P = P_1 + P_2 - \frac{t}{n_t} P_1 P_2$$

where t is the separation of the lenses and n_t is the index of refraction of the separating medium.

*An aperture stop is a diaphragm with an aperture centered on the lens axis. It is used in many camera lenses to control aberrations. The aperture diameter controls the diameter of the imaging pencil of rays and the axial location of the stop determines the zone in which an off-axis pencil passes through a particular lens element.

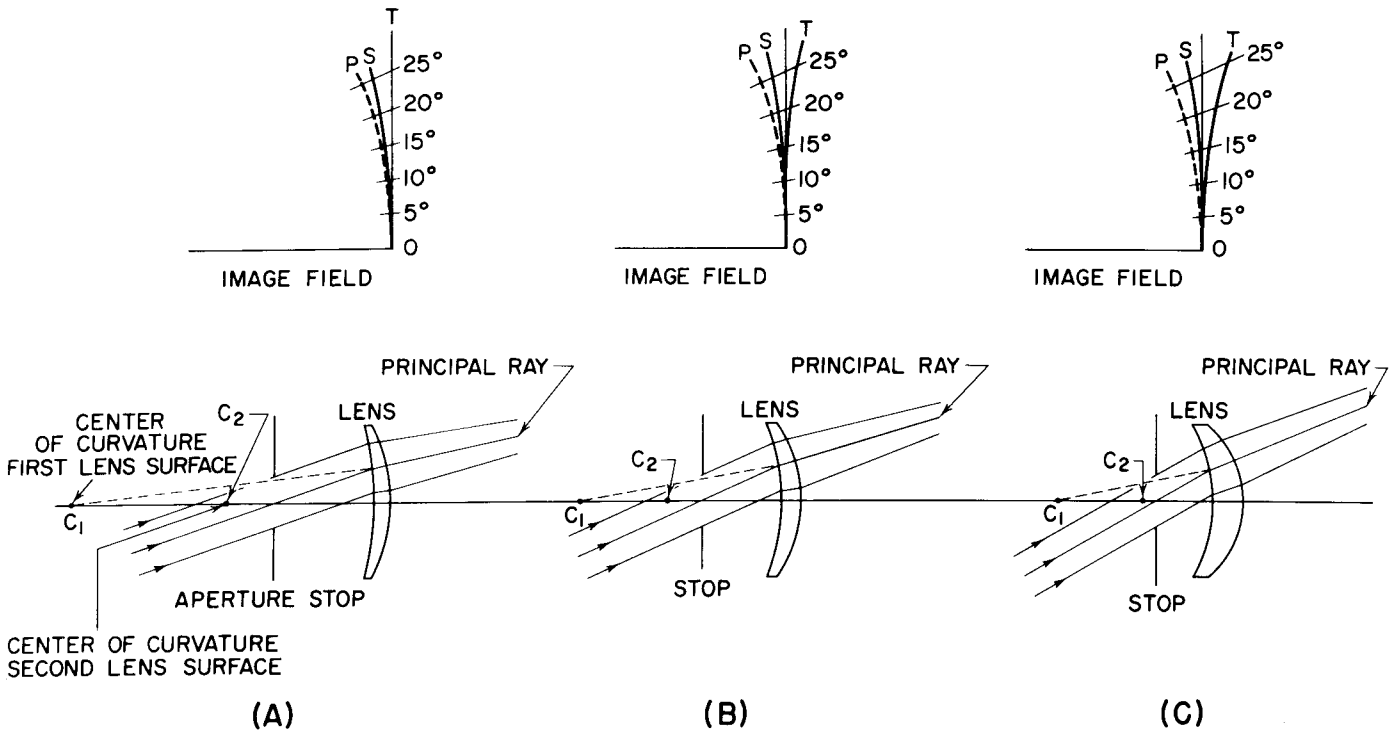


Figure 13. Flattened Image Field Using a Meniscus Lens with Properly Located Aperture Stop

made, most of the region through which the electrons travel has convergent lens action. If electrons cross the convergent equipotential surfaces at other than normal incidence, positive astigmatism is introduced which results in image curvature greater than the Petzval curvature. The curved cathode introduced by Morton helps to start the electrons in a path normal to the equipotential surfaces. In addition, the curved cathode helps to provide uniform curvature in the equipotentials from the axis to the edge of the lens. Although negative astigmatism in part of the path of the electrons can be used to reduce the over-all positive astigmatism, negative astigmatism can only be introduced in a divergent-lens region; however, the divergent-lens region near the anode potential is too weak to introduce significant negative astigmatism.

In summary, where electron trajectories are not normal to equipotential surfaces in the convergent region, positive astigmatism is introduced which cannot be compensated by negative astigmatism from the weak-divergent region of the lens near the anode. Positive astigmatism is accompanied by increased curvature of field.

In the electrostatic lens of the image tube, the location of the crossover point corresponds to the location of the aperture in light optics. The position of the crossover point can be varied by the shape of the cathode and by the distribution of the electrostatic field near the cathode. The center of curvature of the cathode and all equipotential surfaces should be near the end of the anode facing the cathode. If the field between the cathode and the crossover point is radial and directed towards the crossover point, then all equipotential surfaces are spherical, all electron paths are along radii generating the equipotential surfaces, and astigmatism is zero. Schagen, Bruining, and Francken[23] analyzed the imaging properties of such electrostatic-lens systems and derived useful equations for calculating the image position and magnification. However, such an electrostatic lens is not usually achieved in practice because the cylindrical sidewalls of the cathode distort the field from the true spherical shape. With a sufficient number of cylindrical grids of the proper length and diameter, and with the proper applied potential, the spherical shape of the equipotential surfaces may be closely approximated.

In a comparison of the electron lens with an optical-lens system, treatment is given only of spherical refracting surfaces. Equipotential surfaces are reasonably spherical near the axis but may depart from spherical symmetry near the walls of the electrodes. The aspherical shape of the equipotential surfaces may in some cases reduce some of the aberrations but, usually, difficulty is encountered in avoiding those aspherical shapes which increase the aberrations. The cathode sidewall, for example, introduces excess curvature in the equipotential surfaces near the edge which increases astigmatism, curvature of field, and distortion.

Distortion. Linear distortion is present in an image if the lens is not corrected for spherical aberration for the entrance and exit pupils. Pincushion distortion results if the aperture stop (or crossover point) is behind the lens and barrel distortion if the aperture stop is in front of the lens. If the lens is symmetrical about the stop, distortion is absent. Fig. 14 illustrates how distortion arises with the aperture stop in front of the lens.

In image space, a virtual image of the aperture (entrance pupil) appears to be located behind the stop. This virtual image of the entrance pupil is the exit pu-

pil. Because of the spherical aberration of the lens, the exit pupil is farther back if it is imaged by edge rays. A principal ray AL from the object, passing through the center of the entrance pupil, appears in image space to have come from the center of the exit pupil. The image must terminate somewhere along the line PLB; it is assumed that the lens has the proper strength to form the image at IB.

The displacement CL of the principal ray from the axis at the lens, is proportional to the displacement OA of the object point from the axis. If the exit pupil is fixed in position, the displacement IB of the image from the axis is also proportional to the displacement of the principal ray from the axis at the lens. Consequently, in this case, the magnification of the lens is constant over the entire image and the lens is said to be free of distortion.

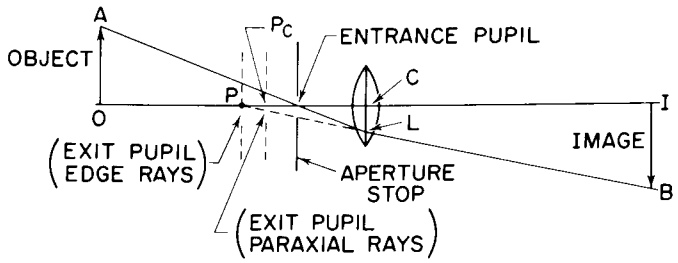


Figure 14. Occurrence of Distortion with Aperture in Front of Lens

However, if the exit pupil recedes from the lens as the displacement of the principal ray from the axis increases, the ratio, (image displacement from axis)/(principal ray displacement from axis at lens) decreases for large displacements. In this case the magnification of the lens (image size)/(object size) is less in the outer zones of the image than in the central zones. The lens then is said to have barrel distortion.

In electrostatically focused image tubes, the situation is similar, but not identical, to the case with the aperture behind the lens. The center of the entrance pupil is determined by the point on the axis towards which the electrons appear to converge. In light optics, the entrance pupil is fixed, but, in an image tube, all rays may not be started towards a common crossover point in the early part of the path. Hence the concept of entrance pupil is only applicable in a general sense. A common exit pupil for all rays is desirable to minimize distortion; a spherical cathode is necessary even to approximate this condition. It is easily seen that the conditions necessary to minimize distortion are similar to those required to minimize astigmatism. In some special tube designs distortion has been made almost negligible, but, with the exit pupil following the convergent region of the lens, pincushion distortion is usually encountered.

A magnetic lens frequently introduces rotation of the image. If rotation is not equal in all zones of the image, then the image is afflicted with spiral distortion. In image tubes, rotation is minimized by keeping the magnetic field as parallel as possible with the principal electron trajectories from every point on the cathode.

REFERENCES

1. Holst, G., J.H. DeBoer, M.C. Teves, and C.F. Veenemans, "An Apparatus for the Transformation of Light of Long Wavelength into Light of Short Wavelength," Physica, Vol. 1, pp. 297-305; 1934
2. Morton, G.A., and E.G. Ramberg, "Electron Optics of an Image Tube, "Physics, Vol. 7, No. 12, pp. 451-459; 1936
3. Krizek, V., and V. Vand, "The Development of Infrared Technique in Germany, "Electronic Eng., Vol. 18, pp. 316, 317, 322; 1946
4. Coltman, J.W., "Fluoroscopic Image Brightening by Electronic Means, "Radiology, Vol. 51, pp. 359-67; Sept., 1948
5. Teves, M.C., and T. Tol, "Electronic Intensification of Fluoroscopic Images," Philips Technical Review, Vol. 14, p. 33; 1952
6. Jenkins, J.A., and R.A. Chippendale, "High-Speed Photography by Means of the Image Converter, Philips Technical Review, Vol. 14, No. 8, pp. 213-225; Feb., 1953
7. Stoudenheimer, R.G., and J.C. Moor, "An Image-Converter Tube for High-Speed Photographic Shutter Service," RCA Review, Vol. XVIII, No. 3, pp. 322-331; Sept., 1957
8. Linden, B.R., and P.A. Snell, "Shutter Image Converter Tubes," Proc. IRE, Vol. 45, p. 513; April, 1957
9. Bliss, W.H., and J.E. Ruedy, "An Electron Tube for High-Speed Teleprinting," RCA Review, Vol. XVI, No. 1, pp. 5-15; March, 1955
10. Morton, G.A., J.E. Ruedy, and G.L. Krieger, "The Brightness Intensifier," RCA Review, Vol. IX, No. 3, pp. 419-432; Sept., 1948
11. Stoudenheimer, R.G., J.C. Moor, H.L. Palmer, "Image Intensifiers for Nuclear Track Imaging," IRE Transactions on Nuclear Science, Vol. NS-7, No. 2-3, pp. 136-141, June, 1960
12. Zavoiskii, E.K., M.M. Butslov, A.G. Plakhov, and G.E. Smolkin, "Atomic Energy" (USSR), No. 4, pp. 39-42; 1956
13. Lallemand, A., and M. Duchesne, "Sur un recepteur ideal de photons et sa realisation," Compt. rend. Acad. Sci. (Paris), Vol. 233, p. 305; 1951
14. Sternglass, E.J., and M.M. Wachtel, "Transmission Secondary Electron Multiplication for High-Speed Pulse Counting, "IRE Transactions on Nuclear Science, Vol. NS-3, No. 4, pp. 29-32; Nov., 1956
15. Prescott, C.H., Jr., and M.J. Kelly, "Cesium, Oxygen, Silver Photoelectric Cell," Bell System Tech. Jour., Vol. 11, p. 334; July, 1932
16. Iams, H., G.A. Morton, and V.K. Zworykin, "The Image Iconoscope," Proc. IRE, Vol. 27, p. 544; Sept. 1939
17. Maloff, I.G., and D.W. Epstein, "Theory of Electron Gun, "Proc. IRE, Vol. 22, No. 12, pp. 1386-1411; Dec., 1934
18. Spangenberg, K., and L.M. Field, "Some Simplified Methods of Determining the Optical Characteristics of Electron Lenses," Proc. IRE, Vol. 3, pp. 138-144; March, 1942

19. Gans, R., "Electron Paths in Electron Optics," Z. Techn. Physik, Vol. 18, No. 2, pp. 41-48; 1937
20. Zworykin, V.K., G.A. Morton, Television, pp. 106-110, John Wiley and Sons; 1940
21. Henneberg, W., and A. Recknagel, "Der Chromatische Fehler bei Elektronenoptischen Anordnungen Insbesonder beim Bildwandler," Z. Techn. Physik, Vol. 16, No. 8, p. 230; 1935
22. Conrady, A.E., Applied Optics and Optical Design, pp. 287-299, Oxford University Press; 1943
23. Schagen, P., H. Bruining, and J.C. Francken, "A Simple Electrostatic Electron-Optical System with only One Voltage," Philips Res. Rep., Vol. 7, pp. 119-130; 1952

Design and Processing of Storage Tubes

M. D. Harsh

Lancaster

STORAGE TUBES AND INFORMATION PROCESSING

Data processing is a term that should be familiar to all who follow the current literature in the field of electronics. Vast sums of money are being invested in research, design, and production activities on a wide variety of often extremely complex and sophisticated electronic machines that perform the function of data or information processing. Computers, radar systems, weapon countermeasures systems, even systems which communicate with space vehicles are some machines in this category, to cite a few examples. Most of these machines have one or more component units used for information storage. This storage unit may be a punched card file, a magnetic tape recorder, a photographic film recorder, or a storage tube, among other examples.

Information storage implies the reception of a message (usually in the form of a sequence of electrical signals), retention of this message for a period of time which may vary over a wide range, and the extraction of this message on command for further processing or for display in a meaningful form. In some instances, the function of a storage device transcends the relatively simple operation of information storage in that distinct manipulations of the data being processed are also performed. For example, the Radechon, one kind of electrostatic storage device, may be operated so that it automatically and internally rejects repetitive elements in a sequence of signals. Another kind of electrostatic storage device, the display storage tube, internally converts the energy of an electrical message to visual form, and additionally performs a storage function in the sense that the message may be rendered visible for periods of time substantially in excess of its original time duration.

This article is concerned only with the electrostatic-charge type of storage tubes. A wide variety of storage tubes for an even wider variety of applications has been proposed. [1] It is possible to make general comparisons relating the advantages of electrostatic-charge storage tubes to the advantages of other information storage units only when the comparison is oriented to specific applications. For example, the amount of information available per unit area on the viewing screen of a display storage tube is much less than that available on a photograph. When the relative speeds of these two media are compared in terms of the time consumed between data acquisition and display, however, the display storage tube has an overwhelming advantage. Generalizations, therefore, must be limited to the following comparative comments. As a rule, electrostatic-charge storage tubes are preferable for information storage in situations where speed of operation, minimum power consumption, and ease of integration with other electronic elements are of prime concern. On the other hand, systems in which gross data capacity and permanence of storage are of major importance require the use of other types of storage media. Occasionally, functional requirements of a system can be met only by exploiting the unique properties of some of the specialized electrostatic storage tubes. Comparisons with other storage media in these cases are, of course, irrelevant.

Electrostatic-charge storage tubes may be classified into three kinds, according to the form of the energy input to the tube and the energy output from it.

ELECTRICAL/ELECTRICAL storage tubes are those in which both input and output are in the form of electrical signals. Tubes in this class typically perform the functions of signal delay, signal integration, time-base conversion, and/or automatic signal comparison.

ELECTRICAL/VISUAL storage tubes accept electrical signals and convert them into meaningful patterns of light. These tubes are useful in the display of information which occurs too rapidly or too slowly to be readily perceived on the screens of conventional cathode-ray display devices. They are also capable of providing readily visible displays in the presence of high ambient illumination.

VISUAL/ELECTRICAL, a third class of storage-type tubes, will not be considered in this article. These devices are more conveniently discussed as camera tubes.

RCA has been active since World War II in the design, development, and production of storage tubes. In the section which follows will be found brief descriptions of the application, construction, and operating principles of storage tube types which are now, or recently have been, in production at the Lancaster plant.

DESCRIPTION OF RCA STORAGE TUBES

Oscillograph-Type Electrostatic Storage Tube

The oscillograph-type electrostatic tube is diagram-

med schematically in Fig. 1; it is intended for service as a memory device in large-scale, high-speed digital computers. The application to which this tube is put started with the use of oscillograph-type cathode-ray tubes. One of the first practical demonstrations of a binary digital computer memory bank of this type is attributable to Williams and Kilburn[2] in England in the late 1940's. A feature unique to this system at the time of its first demonstration was the capability of rapid, random access to a relatively large body of stored binary information.

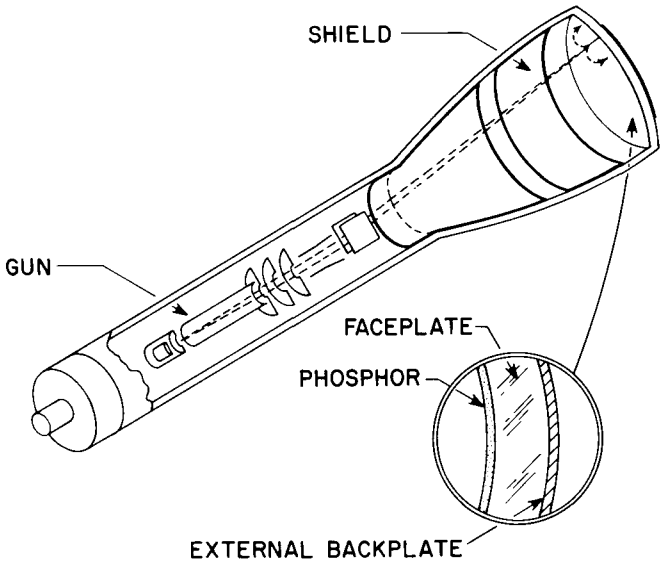


Figure 1. Oscillograph-Type Electrostatic Storage Tube

Superficially, the oscillograph-type electrostatic storage tube is indistinguishable from the conventional cathode-ray tube found in laboratory test instruments. It, too, has an electron gun, a phosphor screen, and a vacuum enclosure or envelope. In its role as a storage device, however, this tube must be fitted with a conducting electrode, which is applied to the external surface of the glass faceplate. This electrode and the glass faceplate with the layer of phosphor deposited upon its inner surface make up the target of this storage tube.

In operation, the electron beam from the gun is positioned in one of N possible locations and turned on by a voltage pulse applied to the control grid. If the beam is turned off while in its original position, the phosphor area under bombardment will be relatively more positive in potential than the immediately surrounding area. The reason for this difference is that the secondary-emission ratio of the phosphor surface is greater than unity and many of the excess electrons have been redistributed in the region immediately surrounding the bombarded area. On the other hand, if the beam is deflected a short distance to the side of its original position before it is turned off, redistribution of electrons will take place into the initially bombarded area, which will then shift in potential in a negative direction. Each of the N possible beam locations on the storage surface plane may be charged to one of two potential levels, depending upon the sequence of control operations performed upon the beam. These two levels may now represent the digits in the binary number system. Interrogation of any one of the N locations on the target is accomplished by positioning the beam on the proper location and then turning it on. The character of the displacement current which flows in the conducting member external to the faceplate depends on the state of charge at the location under interrogation. External circuitry will, of course, interpret the voltage signal developed across a resistance in this circuit in terms significant to the computer.

Radechon

The Radechon, or barrier-grid storage tube,[3,4] shown in Fig. 2, was originally designed as a device for the dynamic differencing of successive signal sequences. As such, it was proposed for use in Moving Target Indication (MTI) radar systems. Subsequently, however, it was recognized that the device possessed considerable versatility and, consequently, a variety of applications were developed in exploitation of its properties. Among these were controllable signal delay, digital data storage, video signal to noise ratio improvement, and time-base conversion (bandwidth compression, for example).

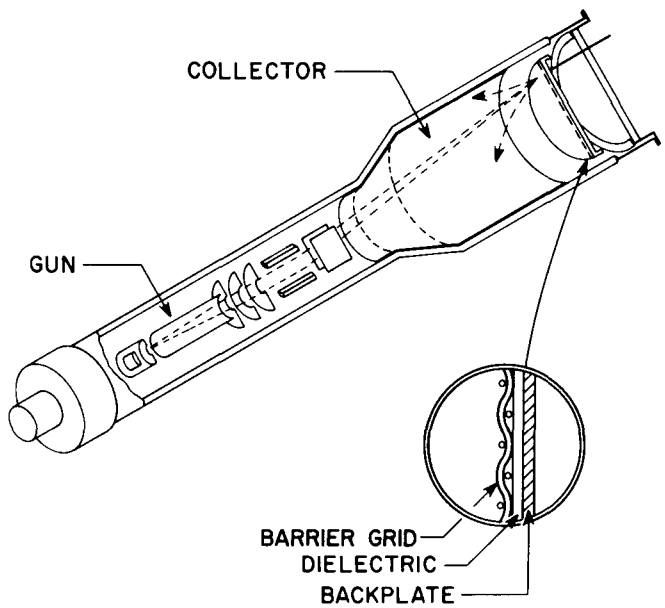


Figure 2. Radechon

The Radechon employs a single gun mounted within a glass envelope on an axis perpendicular to the center of a storage target. The side of the storage target facing the electron gun consists of a fine mesh screen called the barrier grid. Immediately beyond the barrier grid is a layer or sheet of dielectric material. Still farther from the gun, in intimate contact with the dielectric, is a conducting element called the backplate. Another element is a secondary-electron collector in the form of a conducting coating on the inner surface of the bulb wall.

The scope of this article does not permit a detailed description of the several modes of operation possible

with the Radechon storage tube. Therefore, only a very simple sequence involving the manipulation of the charge situation at the Radechon target surface will be described.

The barrier grid is normally operated at reference or ground potential. Choice of the gun cathode potential is dictated by the requirement that the beam arrive at the target with a potential energy insuring greater-than-unity secondary-emission ratio at the insulator surface exposed through the barrier-grid openings.

Assume that the backplate receives a positive pulse of voltage concurrently with linear deflection of the beam across the target. Because of the close capacitive coupling involved, the free surface (the surface nearer the gun) of the insulator is shifted in potential to practically the same extent as is the backplate. If the backplate voltage pulse is of sufficient amplitude, secondary electrons produced as the beam impinges upon the insulator surface cannot escape the retarding field set up by the barrier grid at ground potential. An excess of negative charge now exists along the line of the target scanned by the beam, thus depressing the potential of this portion of the surface. If this process were repeated a sufficient number of times, the surface potential of the insulator surface would eventually be depressed to approximately that of the barrier grid, at which time a potential equilibrium would exist. Under this condition, a potential difference corresponding to the amplitude of the applied backplate pulse would be developed across the target insulator thickness along the line scanned. This process represents writing.

Suppose now that the backplate voltage pulse is not applied during the interval of beam deflection. With the backplate at barrier-grid potential, the potential difference built up across the target insulator thickness along the line scanned previously imposes a negative potential upon the surface of the target in this region. Secondary electrons, produced as the target is scanned, come under the influence of the accelerating field set up between the insulator surface and the barrier grid and are swept through the openings in this grid. These electrons are attracted to the collector. Compensation for the net loss of electrons from the target to the collector is provided by a displacement current which flows in the backplate circuit. This current constitutes an output signal generated as a result of the charge stored in the target during the earlier writing operation. The process just described is called reading.

Because the reading process involves the loss of negative charge, it is likewise an erasing process. If reading is repeated enough times the result is the return of insulator surface potential along the scanned line to barrier-grid potential, thus destroying the gradient previously established. When the equilibrium is thus established, no current will flow in the backplate circuit, and the storage target will be in an erased state.

Graphechon

The Graphechon, [5, 6] as shown in Fig. 3, is intended for service in scan-conversion systems. In one such system, for example, radar video signals are converted in such a way as to be viewable on conventional television picture monitors. Such a system permits the display of radar signals at higher brightness levels than are possible with usual Plan-Position-Indicator (PPI) radar 'scopes. In addition, it provides the electrically controllable persistence of the image, the relatively simple superposition of signals from several sources, and the capability of employing an essentially unlimited number of monitors for remote viewing.

The Graphechon storage tube consists of two electron guns mounted on a common axis on opposite sides of a storage target. The envelope of this tube type serves its classic function as a vacuum enclosure and, to a somewhat more important extent than in other cathode-ray devices, as a support for aiming the electron guns. The target structure of the Graphechon consists of a thin film of insulating material deposited upon a thin conducting membrane which is in turn supported by a fine mesh screen. The membrane and the mesh screen are collectively called the backplate. An essentially annular element, called the anti-shading electrode, is mounted at the periphery of the insulating side of the target. A metal cylinder serves as a collector of secondary electrons emitted from the insulating side of the target; it may additionally serve as the output signal electrode. An envelope wall-coating on the opposite side of the target acts as the ultor for the writing gun.

In normal operation, the collector of the Graphechon is at reference or ground potential. The backplate is operated a few volts below and the anti-shading electrode a few volts above this reference. By virtue of the potential applied to the cathode of the reading gun, the reading-beam bombards the insulating surface of the target at an electron energy resulting in greater-than-unity secondary-emission ratio. Under these circumstances, prolonged scanning of the target surface in a uniform pattern causes the potential of this surface to come to an equilibrium value close to the potential of the collector. The action of the anti-shading electrode is such as to promote uniformity of this equilibrium potential from point to point over the target surface. Under conditions of equilibrium, the current flow in the collector circuit reaches a steady-state value and a potential difference is developed across the thin insulator film.

The function of the writing process is to selectively reduce or eliminate this difference and establish on the free surface of the insulator a potential pattern corresponding to the input signal. The effect of the scanned reading beam in re-establishing the equilibrium thus disrupted is to produce a current flow in the collector circuit. This current constitutes the output signal of the device. Writing is accomplished by causing the input signal to modulate the beam from the writing gun. The beam is scanned concurrently in an appropriate pattern. Wherever this beam strikes the target, it penetrates (by virtue of its acceleration potential) the openings in the mesh screen, the thin conducting membrane, and the insulator itself. Passage of the writing-beam electrons, as well as charge carriers induced within the body of the nominally insulating material itself, perform the discharging function desired. The bombardment-induced conductivity effect just referred to is usu-

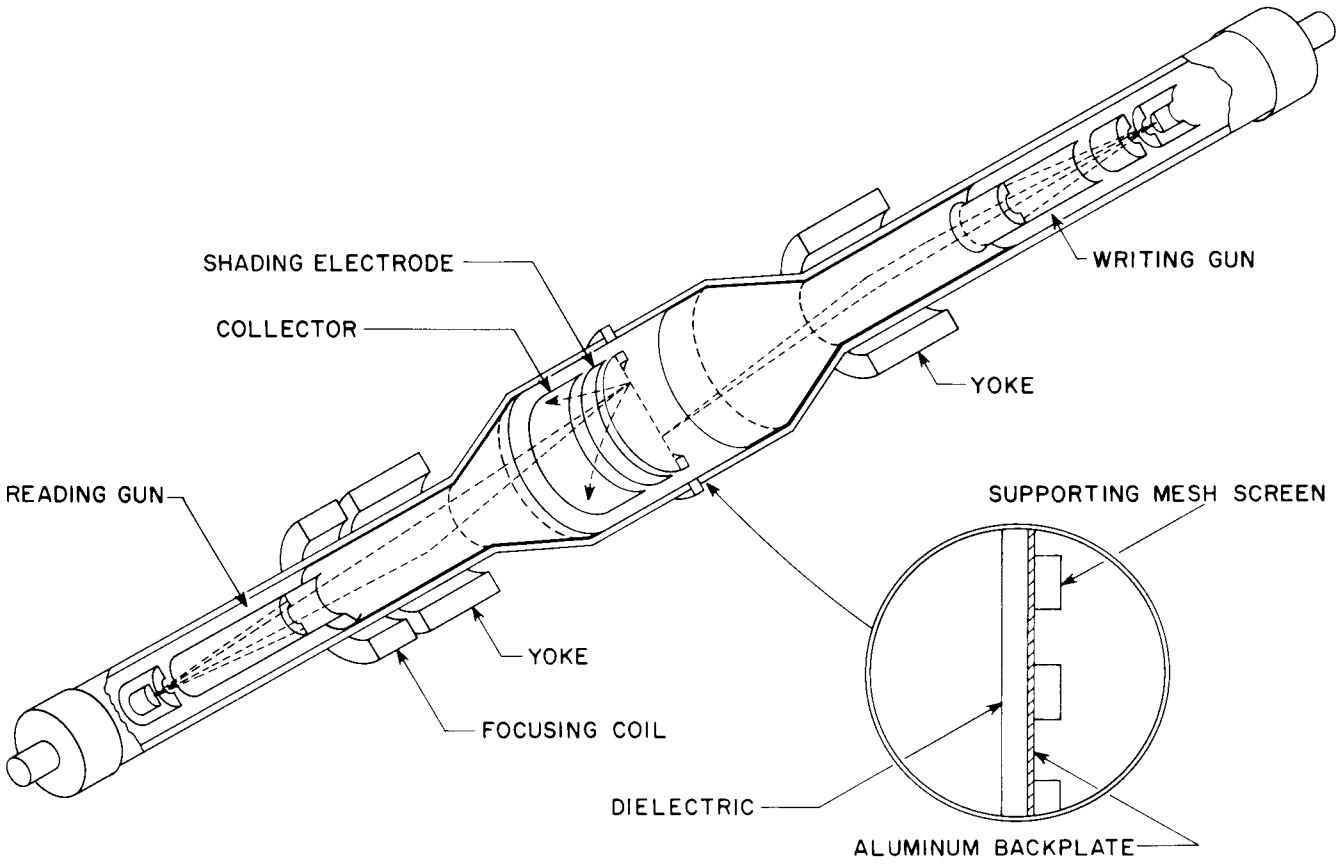


Figure 3. Graphechon

ally the predominant factor in the writing process. Maximizing the effect and, hence, the writing-speed capability of the Graphechon requires that the writing-gun accelerating potential be chosen so that the beam energy is almost completely absorbed in its passage through the insulating film.

Display Storage Tube

The display storage tube[7,8] is used to convert electrical signals directly into patterns of light. In contrast to conventional cathode-ray tubes which perform the same general function, the display storage tube has a very high light output capability and controllable persistence. These tubes are useful in such applications as radar display indicators for use in aircraft cockpits, fire-control display indicators, sonar display units, slow-speed video-data transmission monitors, and oscilloscopes for the study of both high- and low-speed transient phenomena.

Among the most important functional elements of a display storage tube (see Fig. 4) are an aluminized phosphor screen on the inner surface of the faceplate of the envelope, a storage target in the form of a fine metallic mesh screen coated with a dielectric layer on the side away from the phosphor screen, and a second metallic mesh screen called the collector. These elements are all mounted perpendicular to the major axis of the envelope and are arranged with relatively close spacing in the order mentioned. On the axis of the envelope a few inches from the collector is a simple electron gun, whose function is to flood the storage target, through the collector, with a beam of low-velocity electrons of uniform density. Painted upon the inner surfaces of the walls of the envelope, or else supported closely thereto, are conducting surfaces (cylindrical, conical, or hemispherical - depending upon the specific tube design) to which potentials are applied which cause the flooding electron beam to approach the target perpendicular to its surface. Mounted off the tube axis, in some cases in a separate neck, is an electron gun of conventional design which functions as the writing gun. In some versions of the display storage tube, still another off-axis gun is provided for erasing stored information.

In the operation of display storage tubes, the signals to be displayed are applied to the control grid of the writing gun. Concurrently, the resulting modulated beam is deflected in a pattern appropriate to the input signal. This beam is focused so as to impinge upon the dielectric layer of the storage target with the smallest possible cross section. The potential of the cathode of the writing gun is chosen so that the beam will arrive at the target with a potential energy between those of the secondary-emission cross-overs associated with the target dielectric surface. Because the collector is operated at a voltage more positive than that of the target surface, the effect of the writing-beam bombardment is to establish a pattern of varying positive potential which corresponds to the input signal.

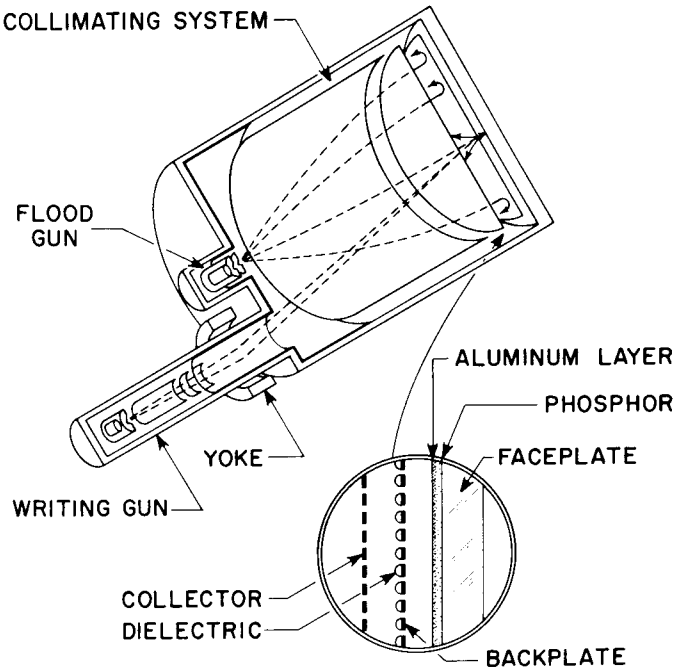


Figure 4. Display Storage Tube

The flooding electron beam is normally operated continuously. In the absence of input signals to the device, steps are taken to insure that the potential of the storage-target insulating surface is sufficiently negative so that no electrons originating at the flood-gun cathode can penetrate the openings in the target. Wherever the potential of the insulating surface is made less negative as a result of writing-beam action, however, penetration of the mesh openings by the flooding-beam electrons can occur. These electrons are accelerated to the phosphor screen by an applied potential and thus produce a pattern of light corresponding in geometry and proportional in intensity to the pattern of potential on the storage-mesh insulating surface.

The stored pattern, which otherwise would persist, may be erased by returning the potential of the insulating surface of the target to a negative level which will interrupt or cut off the flow of flood-beam electrons to the phosphor screen. Techniques for accomplishing this, either instantly or gradually, will not be discussed here; but, briefly, they involve the landing on the insulating surface of electrons having insufficient potential energy to cause secondary emission at greater than unity yield. In this way, a negative charge accumulates and depresses the surface potential as desired. Electrons for this function may come either from the flood gun or from a focused and deflected erasing gun provided expressly for this purpose.

THE STORAGE TARGET

The storage target of a charge storage tube is that member on which is established a pattern of charges that is related to the input information introduced into the tube and that controls the generation of the tube's output. The storage target determines the performance characteristics as well as the very nature of a particular tube type. It should be evident from the preceding section that storage targets take physical forms that differ from one category of tubes to another. In this article, this fact will be recognized and, therefore, no all-inclusive compendium of design rules for storage targets will be attempted. Rather, the discussion will cover the choice and processing of materials suitable for target fabrication in a general way.

Storage targets always consist of a layer of a dielectric material deposited on or otherwise closely attached to a metallic substrate or backplate. This backplate sometimes takes the form of a continuous surface; in other cases, it may be a perforated structure such as a mesh screen.

The Backplate - Perforated Type

Backplates of the perforated type (mesh backplates) are used principally in storage tubes in which a stream of electrons is to be modulated in intensity as it travels to the output electrode. Among storage tubes of this type is the display storage tube. In such tubes, the dielectric storage surface itself assumes the perforated nature of the supporting mesh.

Types of Mesh. Electroformed mesh screens are widely used in storage-tube practice. Mesh screens of this type are preferred for use as storage-target backplates because of their very uniform mesh and for the essentially planar surface they provide for the deposition of the dielectric material. These screens are made by electroplating metals such as nickel and copper on suitable masters. Two types of mesh masters are in common usage. RCA's experience in mesh production is based mainly on the use of glass masters.[9,10] With this method, a thin film of an acid-resisting wax, coated on a polished glass plate, is ruled through by a sharply pointed diamond tool on a ruling engine especially designed for this function. Next, the glass plate is dipped into an acid bath and then the glass surface exposed by the ruling is etched to a depth of about 50 microinches. After the acid-resisting wax is removed by a solvent, the surface of the glass bearing the etched grooves is coated with a suitable conductor by a sputtering process. Careful rubbing of the coated surface removes the conductor except for that adhering to the bottoms of the grooves. The remaining conductor serves as the base on which the subsequent plating proceeds.

Mesh screen made by plating from a metal master is purchased from vendors with recognized competence in this field.[11] The metal master is made by coating a metal plate with a photosensitive nonconductor. This photosensitive surface is exposed to light through a ruled glass master having the grooves filled with an opaque material. A developing solution is used to harden the exposed nonconductor and then the unhardened, unexposed material is washed away. The master is now ready for use in the production of mesh by electroplating.

Both types of masters have been used in routine production to make mesh screens with counts as high as 1000 lines per linear inch. Mesh of even higher count is technically feasible and will undoubtedly become readily available as the art advances. Table I lists typical combinations of the parameters which define

the mechanical characteristics of electroformed mesh screens.

Table I

Mechanical Parameters of Typical Electroformed Mesh Screens

Mesh Count (Lines/Inch)	Transmission (%)	Mesh Thickness (Inches)
1000	45 - 50	0.0002 - 0.0001
750	55 - 65	0.0003 - 0.0002
500	25 - 75	0.0007 - 0.0002
250	35 - 80	0.0012 - 0.0003

The use of etched mesh may be contemplated in the design of storage targets for display storage tubes of large diameter. With large tubes, the relatively low counts attainable in etching technology are not necessarily restrictive so far as over-all tube resolution is concerned. Etched mesh is made by the same process[11] used in the fabrication of shadow masks for color kinescopes and will not be outlined here. However, a point worth noting in connection with this type of mesh screen is that the choice of basic material is not as restricted as in the case of electroformed mesh screen. This fact may enter into considerations of structural strength as well as of cost.

Although woven mesh is generally unsuited to storage-target backplate service owing to the aplanar nature of the mesh boundaries, it will be mentioned here because it is used in other storage-tube technology to be discussed later. Woven mesh is fabricated on looms similar to those used in the textile industry. Stainless steel or bronze wires serve as the basic thread. A desirable feature of such mesh screen is its strength in tension. Its tendency to exhibit bunching of the individual wires is undesirable. Bunching leads to distracting disturbances in the output signal from devices in which a woven mesh screen is employed. Mesh counts of 400 lines per linear inch are available, although 230 lines per linear inch is used most widely.

Mounting of Mesh. Mesh screens for storage-tube targets must, in practically all cases, be mounted under tension to assure a flat, wrinkle-free surface. The development of suitable mounting methods must take into account the tension-relaxing nature of tube processing encountered subsequent to mesh mounting, and the end use of the finished storage tube must also be considered. Operation under conditions of shock and vibration require that the mesh screens of a storage tube have no high-Q mechanical resonances in the frequency range up to several hundred cycles per second.

In one scheme of mesh mounting, copper mesh is welded to a suitably substantial annular ring of cold-rolled steel. To prevent tearing the fragile mesh screen during welding, it is sandwiched between the supporting annulus and a relatively flimsy protective ring. During this process the mesh screen is lightly stretched or smoothed by hand to remove obvious wrinkles. A subsequent firing of the mounted screen at 750 C for 2 minutes in an atmosphere of line hydrogen leaves the mesh structure in a taut condition. The reason for this result is the subject of some disagreement. Some say a recrystallization resulting in greater mass density is responsible. Others believe that surface tension effects occurring while the mesh is highly ductile during heating are a more logical cause. This technique is applied most successfully to targets less than three inches in diameter.

"Hot-blocking" has worked well in the mounting of nickel mesh screens on rings up to 15 inches in diameter. In this case, the screen is heated by radiation from electric heating elements, causing the nickel to expand. While in this state the screen is welded to a type-430 stainless-steel annulus at approximately room temperature. On cooling, the mesh structure is then subjected to considerable tension as desired. Stainless-steel woven mesh screens may also be mounted in this way. Both type-430 stainless-steel and cold-rolled steel annuli can be used as supports for this type of screen.

Stretching a nickel mesh screen and its simultaneous attachment to a supporting ring may be accomplished by brazing. A type-430 stainless-steel ring having one plane surface coated with a brazing compound (e. g., BT solder) is laid, prepared surface down, upon a piece of mesh loosely supported from its edges. The ring's weight, sometimes augmented by additional mass, causes the mesh nickel as it expands under the influence of heat to stretch smoothly under the ring. Heating is provided in a hydrogen furnace at 800 C for 10 minutes. After the assembly is brazed, it is removed from the furnace to cool; the mesh, when cool, has acquired the desired state of tension.

Some display storage tubes using spherically curved backplates have been made experimentally. In this instance, the mesh screen to be formed is placed between two thin (approximately 0.06-inch thick) aluminum plates and the whole is shaped between a punch and die on a punch press. After forming, the mesh is carefully trimmed to size and welded on a suitable stainless-steel or cold-rolled steel annulus.

Throughout the preparation of backplates for storage targets, extreme care must be used to avoid soiling the mesh surfaces. Perspiration stains, for example, may adversely affect the electrical properties of the dielectric material subsequently deposited on the backplate. Even very minute dust particles are to be avoided because they may later become detached, causing the overlying dielectric surface to separate from the mesh structure. Generally speaking, washing of finished backplate assemblies is not effective in this connection, because liquid baths rapidly become contaminated. Occasionally vendors supply nickel mesh having visually undetectable localized occlusions of foreign material which become evident in a finished tube by their effect on the electrical properites of the deposited dielectric material. Firing of backplates made from such mesh in air at 425 C for 20 minutes prior to dielectric deposition is used to improve target performance in finished tubes.

The Backplate – Continuous, Electron-Transparent Type

The principles which govern the operation of double-

ended storage tubes of the Graphechon type require that the storage-target backplate be continuous, yet it must not essentially impede the passage of electrons from the writing gun. This requirement implies the use of a backplate of a relatively low-density metal, such as aluminum, having a thickness no greater than about 1500 angstroms. A film of this character, stretched over a supporting ring up to 2 inches in diameter would have little mechanical strength and low resistance to melting under electron-bombardment heating effects. Accordingly, such backplates are supported on mesh screens of the type discussed in the preceding section. In this case, the screen serves solely as a structural member and heat sink.

Mesh screens for this purpose are mounted and then tightened by the firing technique. They are dipped into an amyl acetate solution of Parlodian and spun dry. The Parlodian fills up the openings and thus serves as a temporary substrate for the deposition of the aluminum backplate itself. Deposition of the aluminum is accomplished by a vacuum evaporation process.[12] Subsequent heating of the tube causes the Parlodian film to be driven off, leaving behind the desired backplate of continuous aluminum.

In an alternate method of forming the Parlodian film, the mounted mesh screen is placed in a container of water. A drop of Parlodian solution is deposited on the water surface, where it rapidly forms a film. The mesh is carefully withdrawn from the water (or the water bled from the container) leaving the Parlodian film stretched over the mesh surface. This film, after drying, is used as a substrate for the evaporated aluminum as in the previous method. This technique leads to a thinner substrate for the evaporation process than does the dipping technique. However, it is less suitable as a production method because the thin film is quite prone to pin holes.

The Backplate—Continuous, Electron-Opaque Type

In several types of storage tubes, the electron beam does not impinge upon the backplate of the storage target; this circumstance simplifies backplate design considerations. There are two cases of interest in this respect: the Radechon and the oscillograph-type electrostatic storage tube.

In the Radechon, the dielectric portion of the storage target is usually a sheet of mica, or possibly ceramic, sandwiched between the backplate and a barrier grid. In order that the backplate have continuous intimate contact with the dielectric, as required for electrical reasons, the backplate is formed by sputtering or evaporating a highly conductive metallic film directly upon one surface of the dielectric sheet. This coated surface is then mounted against a more massive disk or cup which provides a means of electric contact and mechanical support.

Radechons have also been made in which the dielectric material is formed as an oxide directly on the backplate. In this case, the material for the backplate is chosen from among those which may be readily oxidized by anodizing or by other suitable methods. For this process, aluminum is a likely candidate.

In the case of the oscillograph-type electrostatic storage tube, the backplate is mounted on the external surface of the bulb faceplate and normally is provided by the tube user. In this case, the backplate may be a metal foil or mesh screen cemented to the faceplate, a film of conducting paint, or possibly a transparent conducting film of tin oxide.

Choice of the Dielectric

The rules which govern the choice of a dielectric material for use in the target of a storage tube vary in detail with the type of tube in question. Nevertheless, certain general considerations are valid in practically all cases; a discussion of these follows and includes workability, stability, volume resistivity, dielectric constant, dielectric strength, and secondary-emission ratio.

Workability. It is immediately obvious from the earlier descriptions of practical storage tubes that dielectric materials which cannot be formed in thin sheets or films are poor candidates for target use. As a general rule, the dielectric layer of a storage target should have a thickness that is small compared with the diameter of the focused electron beam which deposits the charge pattern. Otherwise, excessive capacitative or conductive coupling from one charged element to its neighbors will adversely affect the resolution capabilities of the target. In a later section, attention will be given to methods of forming dielectric layers.

Stability. To be useful for storage-tube targets, the dielectric must not deteriorate under the influence of environments encountered during tube processing or operation. Examples of such environments follow. Heating of the dielectric in the presence of humid air may occur during the closure operation. Exhaust bake-out schedules may include heating to temperatures up to 450 C. Localized heating due to electron bombardment is inevitable in some modes of tube operation. Mechanical stresses attendant upon shock and vibration environments common to many storage-tube applications may detach the dielectric from its substrate.

A special problem applying to the Graphechon is that its proper operation requires the writing electron beam to penetrate the dielectric layer. Materials for Graphechon storage targets must not be prone to the formation of permanent crystal lattice defects, lest the dielectric properties be altered. Materials having homopolar bonding forces (e.g., zinc sulphide) are preferable in this case to those having ionic bonding (e.g., magnesium fluoride).

Volume Resistivity. High volume resistivity is required of storage-target dielectrics to avoid leakage-induced deterioration of the information content of stored-charge patterns. As a matter of fact, other effects (such as spurious redistribution of secondary electrons and landing of gas ions) are usually the primary contributors to charge-pattern deterioration. Nevertheless, the following rule of thumb may be used

to judge the suitability of a dielectric for application to storage targets:

$$\rho \text{ min} = 2\frac{T}{k}\,10^{14}$$

where ρ min = minimum usable volume resistivity in ohms-centimeter
k = dielectric constant
T = In display storage tubes, required viewing duration in seconds
= In Radechons, time between writing and reading operations in seconds
= In Graphechons, frame period associated with the reading process in seconds

Unfortunately, essentially no work relating to the volume resistivity of materials for storage targets has been reported in the literature. The empirical fact that white or colorless crystalline inorganic compounds often exhibit volume resistivities of the order required in storage-tube practice may be used as a guide.

Once again, the Graphechon must be considered as a special case. The rule of thumb introduced above defines the required nominal resistivity for a Graphechon storage-target dielectric. In addition, these materials must exhibit an effect called bombardment-induced conductivity, [13] in which penetration of the target material by the writing beam leads to the release of additional charge carriers within the bulk of the dielectric. The resulting temporary reduction of volume resistivity enhances the writing speed capabilities of the tube and, more importantly, extends the period of time that written information is visible in the output display. Most dielectrics exhibit bombardment-induced conductivity, the magnitude and time constant of which may be altered by appropriate doping of the basic material before or during target formation. This process is to some extent an art, and no attempt will be made here to present recipes appropriate to the several applications in which Graphechons may be employed. Zinc sulphide has been found to be the most suitable basic material discovered to date.

Dielectric Constant. The dielectric constant of materials for use in storage targets affects not only the leakage time constant, as implied in the preceding section, but also the writing, erasing, reading, and viewing times of the completed tube. Table II gives some values for dielectric constants as measured for typical materials used in storage-tube practice. It is worth noting that the dielectric constant of a material in film form is usually lower than that measured on a large crystal of the same material. This phenomenon appears to be related to the fact that thin films tend to be microcrystalline in nature, with interstitial voids which lower the apparent dielectric constant. Fig. 5 shows an empirical relationship between the dielectric constant of a material and its optical index of refraction. This curve may be used for rough estimates of the dielectric constant for a material in the case where only the index of refraction is known.

Dielectric Strength. Despite the relatively low surface potentials which characterize charge-patterns deposited on storage targets, the extreme thinness of useful dielectric films leads to very high potential gradients through the films. Gradients of 10^5 volts per centimeter are frequently encountered. Table III gives dielectric strengths for several dielectrics. The surface condition of the substrate on which a dielectric is deposited is most important in this connection. Gradients in the vicinity of a scratch may be several times the average gradient for the dielectric film as a whole. Scratches and sharp particles adhering to the substrate frequently lead to localized breakdown.

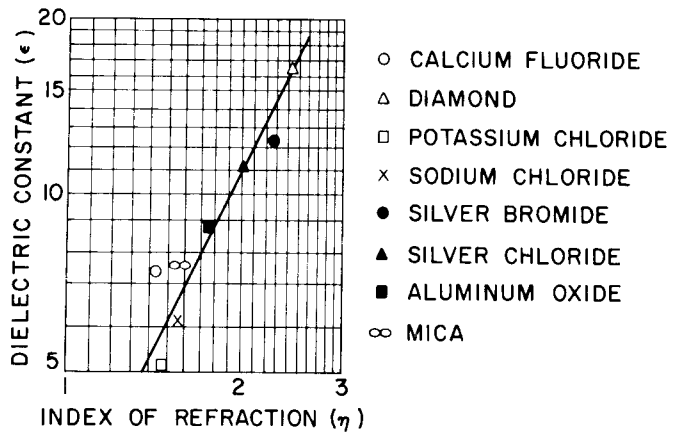


Figure 5. Dielectric Constant Vs. Index of Refraction

Table II

Dielectric Constants for Typical Dielectric Materials

Material	Dielectric Constant
Aluminum Oxide (film)	8.7
Lithium Fluoride	9.0
Magnesium Oxide	9.7
Potassium Bromide	4.9
Sodium Chloride	5.9 - 6.1
Silver Bromide	12.2
Silver Chloride	11.2
Mica	5.4 - 7.6
Calcium Fluoride	7 - 10
Silicon Dioxide (Quartz Crystal)	4.54
Zinc Sulfide (Crystal)	12.1 - 16.5

Secondary-Emission Ratio. The secondary-emission properties of the storage-target dielectric material play an important role in determining the operating parameters of a storage tube. Table IV lists several operating parameters together with the relationships that secondary-emission ratios bear to them. Fig. 6 shows a typical curve of secondary-emission yield as a function of primary-electron beam bombarding potential which may be used to clarify the nomenclature used in Table IV and in the text. As a general rule, δ max and δ min should be, respectively, as large and as small as possible. For a given required writing speed, erasing speed,

Table III

Dielectric Strengths for Typical Dielectric Materials

Material	Dielectric Strength volts/cm
Aluminum Oxide	3.5×10^6
Silicon Dioxide	14.0×10^6
Calcium Fluoride	5.0×10^6
Sodium Fluoride	4.0×10^6
Potassium Bromide	3.0×10^6

or output signal current, therefore, the storage tube may be operated with a minimum value of beam current. This circumstance leads to improved resolution performance and extended tube life. Fig. 7 shows experimentally determined secondary-emission characteristic curves for several useful storage-target dielectrics. The determination of such curves is fraught with many error-introducing pitfalls.[14,15] These data, therefore, should be used as guidance material rather than authoritative statements of incontrovertible fact.

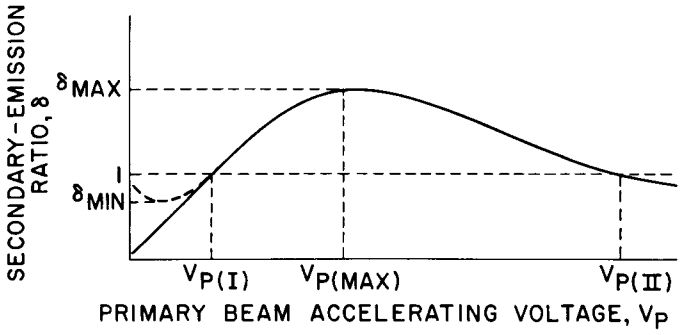


Figure 6. Idealized Secondary-Emission Curve

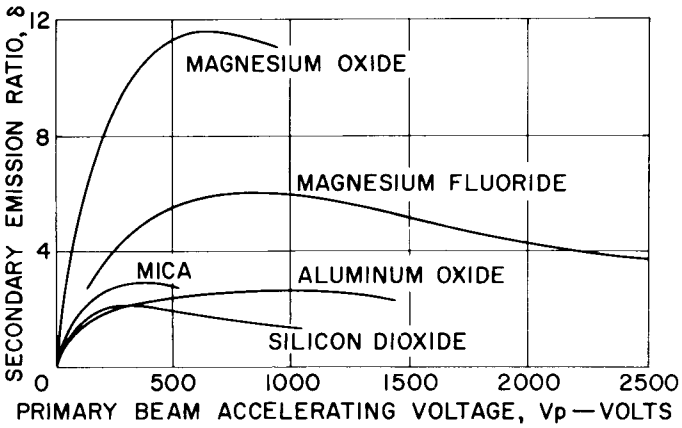


Figure 7. Secondary-Emission Characteristics, Typical Storage-Tube Dielectrics

The Dielectric — Its Preparation

There are a number of methods for preparing thin dielectric layers for storage targets; each method is best suited to a particular category of materials. Typical examples for the several categories will be considered in the following discussion. The greatest attention will be directed toward those techniques of processing which are best known within RCA. Only passing mention will be made of those techniques which are potentially useful but have not yet been seriously developed.

Zinc Sulphide, Magnesium Fluoride, Calcium Fluoride, Cryolite, Lithium Fluoride. Storage targets employing these dielectric materials are invariably prepared by the vacuum evaporation process. Much of the basic technology relating to this method has been developed by workers in the optical field,[16] particularly in connection with anti-reflection lens coatings. It has been a fortunate circumstance that many lens-coating materials display favorable dielectric properties. A brief description of RCA practices in this matter will be given.

Table IV

Relationship of Certain Storage-Tube Operating Parameters to Secondary-Emission Ratio δ

Tube Type	Parameter	Relationship of Parameter to Secondary-Emission Ratio, δ	Potential of Backplate relative to Barrier Grid during writing	Remarks
Display Storage Tube	Writing Speed	$I_b (\delta - 1)$	-------	$V_{p(I)} < V_p < V_{p(II)}$
"	Erasing Speed	$I_b (1 - \delta)$	-------	$V_p < V_{p(I)}$
Radechon	Writing Speed	I_b	positive	
"	"	$I_b (\delta - 1)$	negative	
"	Output Signal	$I_b (\delta - 1)$	positive	
"	"	I_b	negative	
Graphechon	"	$I_b (\delta - 1)$	-------	

I_b = beam current (see Fig. 6 for other symbols)

The substrate to be coated with the dielectric material is supported within a bell jar above the evaporating source. The distance from the source to the surface upon which condensation takes place depends on the dimensions of the surface and the uniformity of thickness required in the resulting film. In the case of two materials used in RCA storage tubes (magnesium fluoride and zinc sulphide) the evaporating source is a boat formed by longitudinally folding a thin strip of platinum. The ends of the boat are clamped securely to two binding posts leading through the base of the vacuum chamber to a low-voltage, high-current source of power. The material to be evaporated is purchased as a chemically pure powder and is compressed into pellets by a pill-making machine. It has been found that pellets splatter less than powder during heating. The pressure in the bell jar is reduced to a maximum of 10^{-4} mm Hg before heating of the platinum boat is started. The current through the boat is increased slowly to permit thorough degassing of the pellet before it is melted and then evaporated. Films in the range of 0.2 to 2.0 microns in thickness are readily prepared by this method.

The thickness of the film of dielectric material condensing upon the metallic supporting substrate is judged by observing the light-transmitting or reflecting properties of a film being simultaneously condensed on a plate of glass mounted in approximately the same plane as the surface being processed. This film acts as an interference filter of continuously changing spectral properties. The color of light from an incandescent source transmitted by (or reflected from) this filter varies through the spectrum as condensation proceeds, the sequence repeating through one or more orders of color until the desired thickness is reached; at this stage, the operator discontinues the evaporation process. Table V indicates, as an example, the color changes observed by reflection as zinc sulphide is condensed on a glass plate. Although this technique has proved to be quite adequate in assuring reproducibility between films deposited in a given setup, more elegant methods are available if greater absolute accuracy is desired. [17] In one such scheme, the incandescent source is replaced by a monochromatic source, and the human observer by a photocell. In this case, the variation of intensity of transmitted or reflected light is detected and related to the thickness of the deposited film.

Considerable art remains in the preparation of thin films by the evaporation method. Both physical and electrical properties of the deposited material may depend upon such factors as the material of the substrate and its surface condition, the rate of evaporation, the number of evaporations performed since the most recent cleaning of auxiliary structures within the bell jar, the presence of volatile substances used in intermediate target-processing stages, and the temperature of the substrate during evaporation, as well as other factors.

Natural Mica. Natural mica is one of the few dielectric materials suitable for use in storage tubes which may be obtained directly in sheet form. Skillful cleavage of the characteristic strata of this material can provide useful layers of less than 0.001-inch thickness. [18] An important defect of mica is ascribable to its natural origin. Physical and electrical properties may vary, depending upon the locale of the mine from which a particular batch was obtained. Mica of quality high enough to be used in storage-tube targets is difficult to obtain in sheets larger than about 25 square inches in area. As a matter of fact, considerable selection is often required to discover circular areas of 2.5-inch diameter which are free from such defects as inclusions of foreign material between strata.

Table V*

Color Changes Observed by Reflection as Typical Dielectric Materials are Condensed on a Glass Plate

Optical Thickness λ = 5461 A	Color Sequence	
	Zinc Sulfide	Cryolite
	Bluish-white	Yellow
$\lambda/4$	White	Magenta
	Greenish-yellow	Blue
$\lambda/2$	Magenta	White
	Blue	Yellow
$3\lambda/4$	Greenish-white	Magenta
	Yellow	Blue
λ	Magenta	Greenish-white

*Source: O.S. Heavens, Ref. 17

A number of methods have been used for cutting mica sheets into the form desired for particular kinds of storage targets. Among these are cutting with scissors, cutting with a razor blade guided by the edge of a metal template, and blanking by a piercing tool and die. One of the more successful methods involves sand blasting. The sheet of mica to be cut is placed between two sheets of high grade, non-abrasive paper. The whole is then placed beneath a metal template on a table, where a sand blast directed at the template can rapidly erode away the exposed mica area.

Aluminum Oxide. As was pointed out earlier in the discussion of backplates, Radechon storage tubes have been made in which the dielectric material is formed directly as an oxide of the backplate material. [19] Anodizing is a convenient process for producing the oxide. Although a variety of metals (titanium, tantalum, tungsten, zirconium, and hafnium)[20] are capable of being anodized, aluminum is usually chosen because of its general availability and the abundance of prior art dealing with its anodizing. [21] The aluminum part to be anodized is used as the anode in an electrolyte containing oxygen-bearing anions (solutions of acids such as sulfuric acid, chromic acid, boric acid, amonium citrate, etc.). When current is passed through the electrolyte, a coating consisting primarily of aluminum oxide is formed on the surface of the part. The thickness of this coating is governed by the voltage applied across the anodizing cell. Typically, a thickness of from 10 to 15 angstroms results for each volt applied. Heating of the

storage target in subsequent tube processing stages may result in crazing and peeling of the coating from the parent metal substrate due to differences in the thermal expansion coefficients of aluminum and aluminum oxide (Al_2O_3). For this reason, the maximum practical coating thickness obtainable is of the order of 0.0005 inches.

Silicon Dioxide. Silicon dioxide films are used in display storage tubes made by another manufacturer. Two methods of fabrication are possible. In one method, silicon monoxide is evaporated onto a metal backplate by the method described above for zinc sulphide. Subsequently, the assembly is heated in air at a temperature of from 400 to 500 C; in this process, the silicon monoxide is transformed to silicon dioxide. In the alternate method,[22] the backplate to be coated is exposed to high temperature (750 to 1100 C) in an atmosphere of ethyl silicate vapor. This vapor decomposes as it strikes the hot backplate to produce a film of silica. Due to the toxicity of ethyl silicate fumes, this method should be used only under carefully controlled conditions.

Phosphors. Storage surfaces consisting of a layer of powder, say of a suitable phosphor, find usage in certain storage tubes. For example, the RCA 6571, an oscillograph-type electrostatic storage tube, uses calcium or magnesium tungstate phosphor material as the storage medium. In this case, the layer is prepared by conventional cathode-ray-tube, screen-settling procedures, modified as necessary to provide exceptional freedom from foreign materials. Storage surfaces of this character have also been prepared by spraying and by dusting.

Miscellaneous Dielectrics. There are several additional materials and techniques for dielectrics which have had only limited consideration within RCA in connection with storage-tube practice. Nevertheless, the storage-tube designer should be aware of their possibilities when he is faced with future design problems.

In this category are sheets of ceramic such as barium titanate, which may be considered as a replacement for mica in Radechon-type tubes. Such sheets may be obtained in about 0.010-inch thickness. Although this material is somewhat difficult to work with due to its brittleness and is somewhat granular in surface condition, these defects may, in the future, be overcome or may, in fact, be unimportant in special design situations.

For storage tubes where a dielectric in the form of a mesh screen is desirable, Corning Fotoform glass should be considered.[23] Fotoform mesh of up to 750-line count is available in experimental quantities.[24] Corning representatives believe that present problems relating to nonuniformity of mechanical and electrical properties will eventually be overcome.

The use of cataphoretic-coating techniques in storage-tube practice[25] may well be considered where relatively thick dielectric layers are required. Thick layers might also be obtained by sintering finely divided glass particles directly to a supporting backplate.

Auxiliary Target-Assembly Elements

In addition to the backplate and the dielectric, a storage-target assembly may include one or more additional elements whose functions may include collection or repulsion of primary or secondary electrons from the dielectric surface, prevention of undesirable redistribution of secondary electrons, establishment of uniform accelerating or retarding fields in the vicinity of the target itself, or other functions which will be evident upon detailed study of the particular type of storage tube.

Elements in this category take the form either of planar or spherically curved mesh screens. The preparation and mounting of fine mesh screens have already been discussed. Reference here, therefore, need only state that the microscopically aplanar nature of woven mesh is frequently found to be desirable for electron optical reasons, particularly in the case of collectors for some display storage tubes. Here the very slight scrambling of the flood or viewing electrons as they pass the plane of the collector minimizes moire effects resulting from the proximity of the mesh-screen collector to the mesh-screen backplate. Woven mesh screens are also often used for the barrier grids of Radechons due to the relatively large thickness for a given electron transmission. In the case of the barrier grid, it is desirable that the secondary-emission ratio of its surface be very close to unity. This effect is accomplished by coating the barrier grid with a thin layer of gold plating, or by evaporation or sputtering techniques.[26]

Some elements in the auxiliary category may be cylindrical sections mounted with their axes perpendicular to the plane of the target itself. These sections may be formed by standard metal-working techniques such as machining, deep-drawing, or wrapping on suitable mandrels. Sometimes, it is convenient to form them directly on the inside of the glass envelope by painting the inner wall with aquadag or silver paste, followed by an appropriate thermal baking. Vapor-coating practices have also been used on occasion to achieve the same end result. In either case, the coatings should be smooth, hard, and essentially free from surface oxides.

Luminescent Screens

Strictly speaking, the luminescent screens of display storage tubes are not considered as part of the target assembly; nevertheless, they are introduced briefly at this point for convenience. In general, the technique of depositing luminescent screens is the same as that used in the manufacture of oscillograph tubes. In addition, care is taken to remove all stray phosphor particles from areas other than the faceplate itself. For proper operation of a storage tube, it is required that the phosphor screen be backed up by a conducting film. This film may be either a conventional aluminized backing on the surface facing the electron guns or else a transparent conducting layer of tin oxide deposited directly on the inside of the face plate before the phosphor screen is laid down.

ELECTRON-OPTICAL STRUCTURES IN STORAGE TUBES

Electron Guns

The primary goal of the designer of electron guns for

cathode-ray tubes is to produce an electron-optical structure that will produce, at the phosphor screen, an electron spot of smallest possible cross section and maximum possible current density. The same criteria generally apply to the design of guns for storage tubes. In one important respect, the designer of guns for storage tubes is at a disadvantage with respect to his colleague in the conventional cathode-ray-tube field. The latter is able to take fullest possible advantage of the fact that focused electron spot sizes improve (become smaller) with increasing acceleration voltage and are limited in this respect only by the equipment designer's restrictions on ultor power-supply voltage and deflection power.

In storage-tube practice, gun accelerating voltages are determined by the characteristics of the target to be bombarded. The optimum accelerating voltage depends not only upon the type of material but also upon the function to be performed by the electron beam in question. The most favorable situation in this connection applies to the Graphechon writing gun. Here it is desired to penetrate the storage target and to excite charge carriers within the bulk of the storage medium. It has been found empirically that voltages of the order of 5 to 10 kilovolts are appropriate for this gun.

More restrictive are the cases wherein the desired function of the electron beam is to generate as many secondary electrons at the storage target surface as possible. This condition applies to the Graphechon reading gun, the display-storage-tube writing gun, and the Radechon and oscillograph-type storage tube guns. As may be seen from the curves of secondary-electron emission yield for various practical storage-target dielectrics (see Fig. 7), optimum yield is invariably obtained with an accelerating voltage of about 1000 volts, or even less. In most cases, a compromise favoring improved spot size is reached, in which accelerating voltages as high as 2500 volts may be employed. A truly difficult situation arises in the case of the selective erasing gun employed in certain display storage tubes. Here the over-all acceleration potential difference between gun cathode and the plane to be scanned with a focused beam (the storage target) is only a few volts, 10 volts at most. Because equipment design considerations limit the average potential of the deflecting electrodes (and hence of the ultor) to the order of 100 volts, focusing takes place at a low velocity level. As a result, spot diameters of one-eighth or even one-quarter of an inch are obtained with a beam current of only microamperes.

The flood gun of the display storage tube is similarly limited to a low over-all acceleration potential - in effect, zero volts. However, this requirement is not as restrictive as in the previous case because here the beam is deliberately defocused and intervening elements such as the collector may be set at potentials as high as several hundred volts. It is regularly possible to deliver flood-beam currents of about one milliampere or more to the vicinity of the mesh collector; of this current, several hundred microamperes may be delivered to the phosphor screen when the insulating surface of the storage target is "turned on" by the writing process.

One factor applicable to storage-tube gun design almost completely counteracts the restrictions on over-all accelerating voltage just discussed, at least in a number of cases. Notable among these are the cases of both guns of the Graphechon, as now produced, and of the writing guns in those display storage tubes intended specifically for low-frame rate, video-display service (e.g., sonar display service). Here very low beam currents are required, a beam of one or two microamperes being quite typical. In these cases, very heavy beam masking can be employed, thus restricting beam diameters in the focus and deflection planes. In addition, space-charge spreading of the focused spot is minimized.

In spite of the exceptions mentioned above, the design of guns for storage tubes generally follows the well-established rules of cathode-ray-tube practice.

The choice of deflecting method in storage-tube gun-design practice is almost completely dictated by applications requirements. For storage tubes intended for computer memory service (oscillograph-type electrostatic storage tube, Radechon), electrostatic deflection is mandatory. This use is a result of the desire to be able to address information locations randomly at high rates, say at 100,000 events per second. Magnetic-deflection systems of this speed, operating at an economical power level, are simply not available.

Electrostatic deflection is also chosen in those cases where writing and reading are accomplished by a single gun (as in a Radechon), quite possibly at widely different scanning rates for each function; where more than one sequence of signals must be time-shared on a single gun (as in the display storage tube); or where very slow scanning rates are encountered (also in the display storage tube), as in sonar display systems for example.

Magnetic deflection is quite appropriate to the display of Plan Position Indicator (PPI) radar information. This use is usually accomplished by a single set of deflection coils, mechanically rotated in synchronism with the radar antenna. This type of deflection is employed in some display storage tubes and in the case of the Graphechon writing gun. The beam of the Graphechon reading gun is likewise magnetically deflected; the major application of this tube at the time of its design involved reading at television scanning rates.

The choice of the focusing component for use with storage tubes follows conventional cathode-ray-tube practice. From the point of view of system weight, bulk, and power consumption, electrostatic focusing is preferred.

Collimating Systems

For proper operation of a display storage tube, it is required that the flood-beam electrons approach the surface of the dielectric perpendicularly. The electron-optical action which produces this result is known as collimation. A simple collimating system for use with a planar target is diagrammed in Fig. 8 along with its light-optical analog. Unfortunately, the focal length of such a system is so long relative to its useful diameter

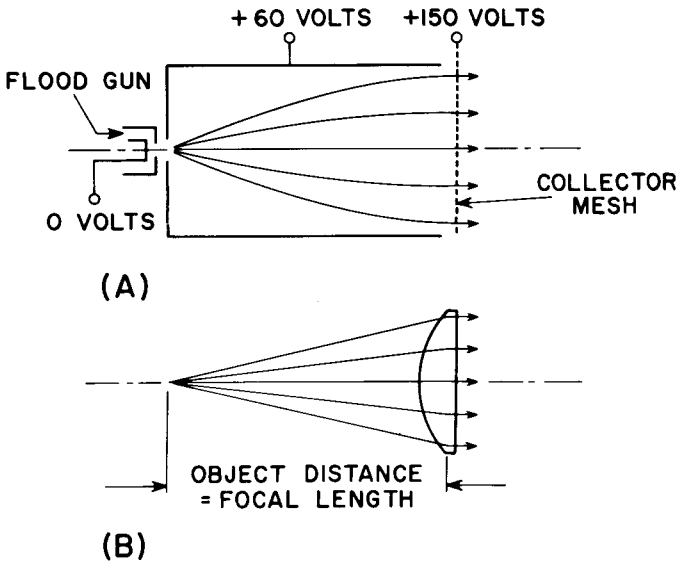


Figure 8. Simplified Collimating System (A) and its Optical Analog (B)

that it is dimensionally unsuitable for use in a practical tube. In addition, spherical aberrations militate against uniform perpendicularity of flood-beam incidence at all target radii. Accordingly, additional elements must be introduced into a practical collimating system.[27] The design of such a system is usually accomplished by a cut-and-try process involving either demountable or sealed-off tubes. Recent advances in the application of resistance-network analogs in conjunction with computer techniques[28] have proved to be most useful in establishing a logical starting point for such cut-and-try experimentation.

Lens Raster Systems

The individual openings in the field configuration around a backplate of a display storage tube give rise to minute electron lenses which determine the paths taken by flood electrons in their flight to the luminescent screen. The totality of an individual lens is described as a lens raster system.[29] The designer concerned with display-storage-tube resolution must give attention to this matter, because mesh-opening dimensions, mesh spacings, and applied voltages have an important bearing on over-all tube performance.[30, 31] As in the case of collimating-system design, substantial insight into the problem is obtained by means of resistance-network analog analysis. Several reports dealing with this technique are available.[32]

CONCLUSION

In the preceding pages, an endeavor has been made to treat briefly the salient features of storage-tube design and processing practice. Of necessity, much pertinent information has been omitted, either for reasons of space or because other articles in this book have covered the subject. Despite the general novelty of storage tubes, many general tube-design rules are apropos. The novice storage-tube designer, therefore, will find profit in a careful study of all of the aspects of tube design.

The author takes this opportunity to acknowledge the many contributions to this article by his colleagues in the storage-tube engineering and manufacturing activities at RCA.

REFERENCES

1. Knoll, M., and B. Kazan, Storage Tubes and Their Basic Principles, John Wiley and Sons, Inc., New York, 1952
2. Williams, F. C., and T. Kilburn, "A Storage System for Use with Binary-Digital Computing Machines," Proc. Inst. Elec. Eng. (London), 96, Part II, pp. 183-202, and Part III, pp. 77-100, 1949, and 97, Part IV, pp. 453-454, 1950
3. Jensen, A. S., "The Radechon, A Barrier Grid Storage Tube," RCA Review, 16, pp. 197-215, 1955
4. Hines, M. E., M. Churney, J. A. McCarthy, "Digital Memory Barrier Grid Storage Tubes," Bell System Technical Journal, 34, pp. 1241-1264, 1955
5. Pensak, L., "The Graphechon - A Picture Storage Tube," RCA Review, 10, pp. 59-73, 1949
6. Dyall, W. T., G. R. Fadner, and M. D. Harsh, "Development of an Improved Graphechon Storage Tube," RCA Review, 13, pp. 546-557, 1952
7. Knoll, M., and P. Rudnick, "Electron Lens Raster Viewing Storage Tube," Washington Symposium on Electron Physics, NBS Circular No. 527, 1951
8. Patrick, N. W., and P. P. Damon, "Recent Developments in Display Storage Tubes," Proceedings of the National Aeronautical Electronics Conference, pp. 201-206, 1959
9. Law, H. B., "A Technique for the Making and Mounting of Fine Mesh Screens," Rev. Sci. Instr., 19, pp. 879-881, 1948
10. Dyall, W. T., "A Ruling Engine at RCA," RCA Engineer, Vol. 2, No. 1, p. 22, 1956
11. Cook, M. C., "Acid Etching and Electroforming Precision Parts," Product Engineering, Vol. 27, No. 7, pp. 194-199, July 1956
12. Strong, J., Procedures in Experimental Physics, Prentice-Hall Inc., New York, pp. 168-187, 1942
13. Pensak, L., "Conductivity Induced by Electron Bombardment in Thin Insulating Films," Phys. Rev., 75, pp. 472-478, 1949
14. Code U1
15. Stambach, G. L., "Secondary Emission Ratio of Storage Tube Insulator Films," NRL Report 5029, 1957
16. Heavens, O. S., and S. O. Smith, "Dielectric Thin Films," J. Opt. Soc., 47, pp. 469-472, 1957
17. Heavens, O. S., Optical Properties of Thin Solid Films, Academic Press, Inc., New York, pp. 242-249, 1955
18. Strong, J., op. cit., pp. 388-389
19. PB149876, Library of Congress, Washington 25, D. C.
20. Misch, R. O., and W. E. Ruther, "The Anodizing of Zirconium and Other Transition Metals in Nitric Acid," Jour. Electro. Chem. Soc., 100, pp. 531-537, 1953
21. Hass, G., "On the Preparation of Hard Oxide Films

with Precisely Controlled Thickness on Evaporated Aluminum Mirrors," J. Opt. Soc. 39, pp. 532-541, 1949
22. Law, H. B., "Formation of Insulating Layers by the Thermal Decomposition of Ethyl Silicate," Rev. Sci. Inst., 20, p. 958, 1949
23. Code U2
24. Byer, M., "Chemical Machining Photosensitive Glass," Materials and Methods, Vol. 43, No. 6, pp. 134-137, 1956
25. Bidgood, E. S., and G. H. Kent, "Cataphoresis and Alundum Coatings," Trans. of Electro-Chem. Soc., 87, pp. 321-329, 1945
26. Strong, J., op. cit., pp. 159-168
27. Kazan, B., and M. Knoll "Viewing Storage Tubes," RCA Industry Service Laboratories Report No. RB-52, pp. 10-12, 1956
28. Code U3
29. Kazan, B., and M. Knoll, op. cit., pp. 7-10
30. Code U4
31. Code U5
32. Code U6

Design and Processing of Microwave Magnetrons

F. E. Vaccaro and D. E. Nelson

Princeton

A magnetron has been defined as a diode immersed in a magnetic field; however, two additions should be made to this definition. First, the dc electric field and magnetic field must be at right angles to each other. The relationship is often expressed by calling the magnetron a crossed-field device. Second, electron movement occurs in such a way that the potential energy due to the dc electric field is converted to rf energy.

The magnetron generally consists of a cylindrical cathode surrounded by a cylindrical anode that may be divided into segments and requires for its operation a magnetic field parallel to the cathode. The rf circuits are connected either between the cathode and anode or between the segments of the anode. The region between the cathode and the anode is called the interaction space. In this region, the electrons interact with the rf field, and the energy gained from the dc field applied between anode and cathode is converted into rf energy.

In the analysis of the operating principles of a magnetron, it is desirable to first consider the case involving only dc operation (no rf voltage applied). An electron emitted from the cathode is acted upon by two radial forces, as shown in Fig. 1. The force, F_1, due to the dc electric field is directed toward the anode and is equal to eE, where e is the charge of the electron and E is the intensity of the electric field. The force, F_2, due to the magnetic field is directed toward the cathode and is equal to $e[v_e \times B]$, where $[v_e \times B]$ is the vector cross-product of electron velocity and magnetic-field density.

When the two forces, F_1 and F_2, are equal and opposite, the electron travels in a circular path around the cathode. In this "equilibrium" condition, therefore, a cloud of electrons circling the cathode extends to a certain radius. Beyond this radius there is no electron space charge. If this radius is less than the anode radius the magnetron is said to be "cutoff", i. e., no electrons can reach the anode. However, if the anode voltage is raised, the radius of the space-charge cloud increases and electrons reach the anode.

A simple, or cyclotron-frequency, magnetron oscillator results if a circuit that is resonant at the cyclotron frequency of the electrons is connected between the anode and cathode of the dc magnetron. The electrons flow from the cathode toward the anode and then back toward the cathode (at the cyclotron frequency). This type of magnetron is very inefficient and is rarely used.

MKS units are used throughout this article except where specified otherwise in the text.

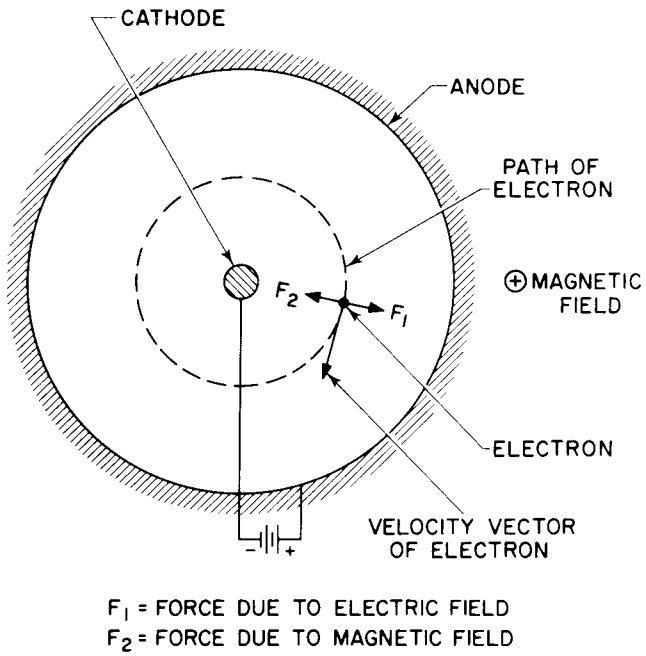


Figure 1. Electron Motion in a DC Magnetron (no rf voltage applied)

Of much greater importance is the traveling-wave type of magnetron. This type is often called a cavity magnetron although it does not necessarily employ cavities, i. e., the circuit could consist of lumped parameter elements or resonant sections of transmission lines. The term cavity magnetron came into use because the first traveling-wave magnetrons of practical importance were made with rf circuits which were resonant cavities.

Fig. 2 shows some typical unstrapped-resonant-systems for traveling-wave types of magnetron oscillators. The anode may be considered as a series of N quarter-wave transmission lines which act as resonators. A system of N coupled resonators can be shown to have N resonant frequencies called "modes," each of which has a different electric field distribution. At these resonant frequencies standing waves can be excited. These

standing waves can be resolved into oppositely directed traveling waves; if the electrons rotate in synchronism with one of these waves, interaction will take place to transfer energy from the electrons to the wave.

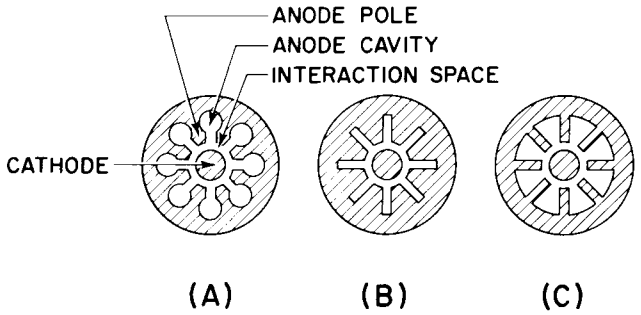


Figure 2. Typical Unstrapped Resonant Systems for Traveling-Wave Magnetron Oscillators: (A) With hole-and-slot-type side resonators, (B) With slot-type side resonators, (C) With vane-type side resonators

Fig. 3 shows a linear form of the traveling-wave magnetron illustrating some of the basic principles. An rf wave is shown traveling to the right. If an electron travels with the same velocity as the waves and in the proper phase, it will continuously lose energy over a number of cycles. The magnetron, therefore, is independent of transit time; i. e., the electrons will continue to lose energy even though their transit time is a number of cycles. If the linear magnetron structure of Fig. 3 is rolled up it assumes the more usual circular reentrant form.

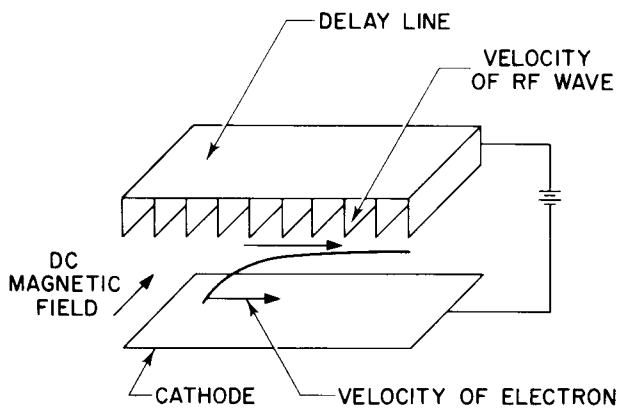


Figure 3. Interaction of Electron with RF Wave in Traveling-Wave Magnetron

Some of the paths of electrons in a traveling-wave magnetron are shown in Fig. 4. The electrons travel in cycloidal paths; however, the design equations ignore the cycloidal motion and deal only with the average paths of the electrons. Electrons out of phase with the radio frequency fields are quickly returned to the cathode. This action is one reason for the high efficiency of the magnetron. A second reason for the high efficiency, mentioned previously, is that the potential energy of the electrons is converted to rf energy so that the velocity remains constant and the electrons, therefore, tend to stay in proper phase. A third reason for high efficiency is the phase-focusing effect which keeps the electrons in the proper phase. Fig. 5 shows a plot of lines of electric force on an electron (drawn for the plane case) for the fundamental of the π mode. The plane of maximum opposing force on the electron intersects that of the figure along the line M. The arrow shown above the line M indicates the direction of electron motion. The force on the electron due to the dc electric field is directed from cathode to anode. The force lines shown may be considered to be those of the total field component at the instant the rf field is maximum. An electron which lags behind the rf field, say at point P, "sees" a radial component of the rf field which adds to the radial electric field divided by the magnetic field; this increased electric field accelerates the electron until it is in the correct phase to give up energy to the field. Similarly, electrons which get ahead of the rf field are slowed down.

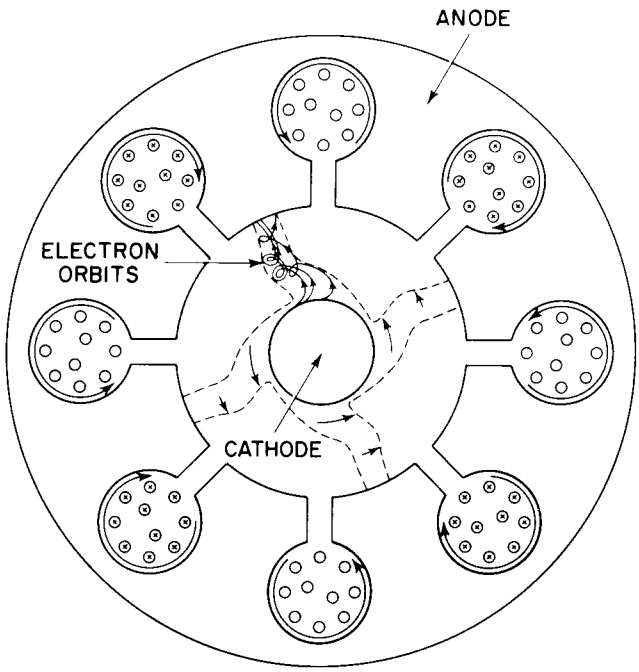


Figure 4. Space Charge in Oscillating Magnetron

It should be realized that the traveling-wave magnetron and the traveling-wave tube are very closely related. Both tubes require a structure to slow down the rf wave so that the electrons can have approximately the same velocity as the rf wave and can continuously lose energy to the rf field. In both types of tubes the electrons may interact with forward waves or backward waves. Both types may serve as amplifiers or oscillators. The basic difference between the two tube types is that, in the case of the magnetron, potential energy is converted to rf energy while, in the case of the traveling-wave tube, kinetic energy is converted to rf energy.

The most important use of the magnetron is as a self-excited oscillator in pulsed radar systems. In the more sophisticated radar systems now being developed, the magnetron oscillator is being replaced by amplifier tubes of the magnetron, traveling-wave, or klystron type. Probably the most important future applications

of magnetrons will be in lightweight radars and in the field of dielectric heating where an efficient source of high-frequency power is all important. At present, magnetrons are being used in electronic ovens and diathermy devices. Other magnetron applications include frequency-modulation communication, weather radar, and radar countermeasures. A typical magnetron for use in pulsed radar is shown in Fig. 6.

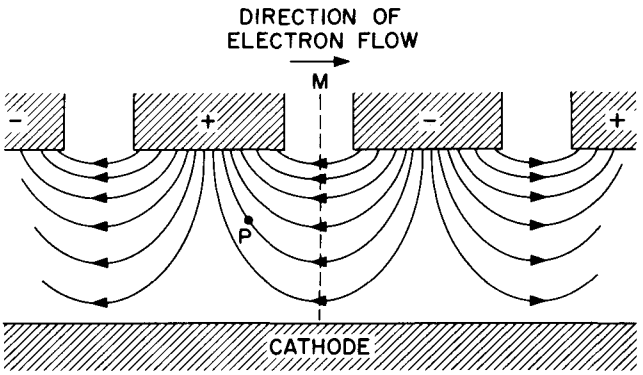


Figure 5. Phase-Focusing Effect in a Traveling-Wave Magnetron

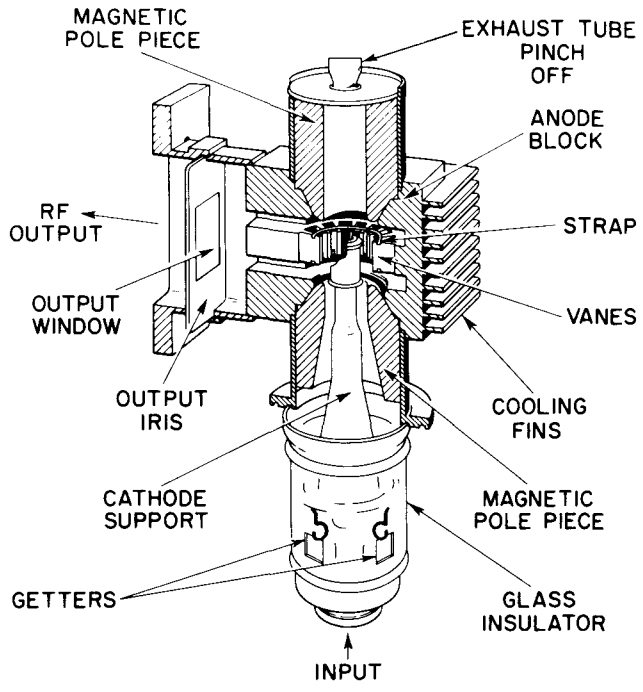


Figure 6. Cutaway View of RCA-6521 Magnetron

DESIGN OF THE MAGNETRON INTERACTION SPACE

Interdependence of Magnetron Components

The components of the magnetron are so interdependent that no one component can be designed independently of the others. The basic problem is, given a set of requisite parameters by the magnetron user, to determine a design for each component meeting these parameters which is also consistent with the design adopted for the other components. Fig. 7 is a chart showing an analysis of the various factors in magnetron design. At the left are the parameters specified by the equipment designer. The relative importance of these parameters may assume equal or greater importance in special applications; for example, frequency stability may be especially important in beacon applications, or weight may be important in airborne equipment.

The four components of the magnetron are indicated in heavily lined blocks, namely, interaction space, resonant system, cathode, and magnetic circuit. The output circuit is considered to be a part of the resonant circuit in the chart although it might have been considered as a separate component. The arrangement of the chart indicates the usual method of magnetron design, i. e., a trial design of the interaction space is determined which satisfies the required parameters. This design establishes a set of specifications for the remaining components. Then it must be decided whether or not each of the remaining components can be designed to meet these specifications. Thus, the specified parameters plus the interaction space design impose sets of conditions upon the remaining elements.

Modes of the Magnetron

In a circular system of N resonators, the phase difference between adjacent resonators may assume values $2\pi n/N$, where n represents the integers 0, 1, 2, ... N/2. The variation of potential from one segment to the next depends upon the mode of oscillation of the whole system of resonators. The mode number n denotes the number of complete waves of potential around the anode. Magnetrons are normally operated in the $n = N/2$ or π mode. These anode potential waves may be either standing waves or waves traveling around the anode structure with angular velocity $2\pi f/n$ radians per second, where f is the frequency of the particular mode n. For the two modes in which adjacent resonators are in phase (when $n = 0$), or π radians out of phase (when $n = N/2$), only standing waves are possible. These modes are a property of the resonant circuit.

An electron traveling around the interaction space can continue to lose energy to one of the modes if it crosses each anode gap in the proper phase. If the time to reach the next gap is $(|q| + 1/2)T$, the electron will cross the next gap when the rf phase is the same as when it crossed the previous gap. The letter q is any integer. It can be shown that this action will occur if the angular velocity of the electron is $2\pi f/|k|$, where $k = n + qN$. The number k is the Hartree mode number and is equal to both the number of rf cycles required for the electron to move once around the interaction space and the number of spokes in the space-charge cloud. If $|k| = n$, the electronic interaction is said to be with the fundamental Hartree component of the rf field. If $|k| \neq n$, the electronic interaction is said to be with a Hartree harmonic. Fig. 8 is a π-mode potential plot showing the Hartree fundamental and several harmonics.

Magnetrons are usually designed for operation in the fundamental component and the π mode. The use of the π mode is indicated by its greater mode separation (for

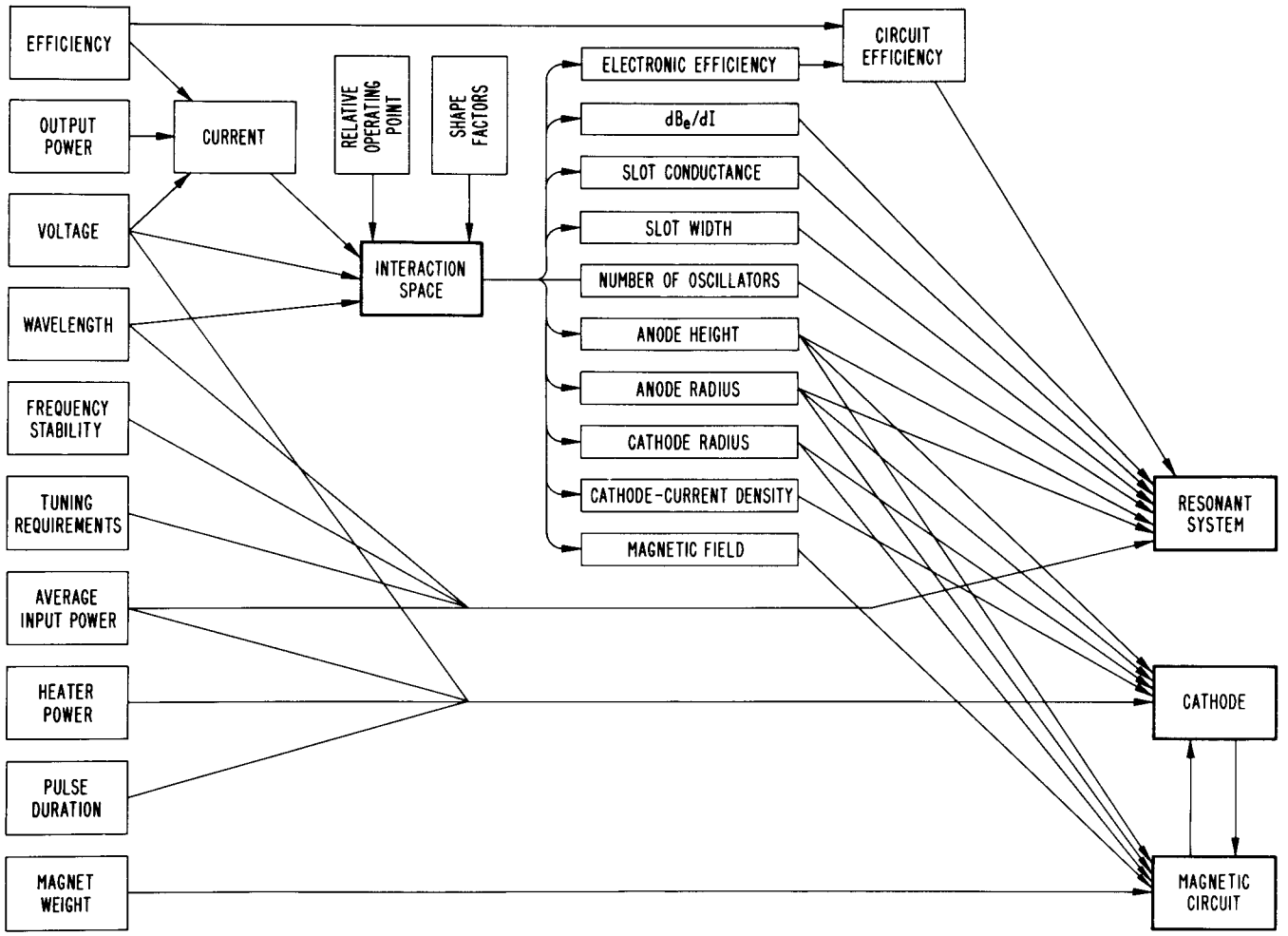


Figure 7. Principal Factors in Analysis of Magnetron Design

a strapped or rising-sun system). The use of the fundamental component gives the most efficient interaction of electrons with the rf field.

The Equations of Cross-Field Interaction

The relationship between the magnetic field B and the voltage at which an electron just grazes the anode of a cylindrical magnetron (with no rf fields present) is given by the Hull cutoff relation:[1]

$$V_c = \frac{eB^2 r_a^2}{8m}\left[1 - \left(\frac{r_c}{r_a}\right)^2\right]^2 \qquad (1)$$

Here e/m is the ratio of electronic charge to mass, and r_a and r_c are the anode and cathode radii, respectively. At voltages below the Hull cutoff voltage V_c no current reaches the anode in the ideal case.

Langmuir's[2] relation for space-charge-limited current flowing between coaxial cylinders in the absence of a magnetic field is:

$$\frac{I}{h} = \frac{8\epsilon_o\sqrt{2}}{9\beta^2}\pi\sqrt{\frac{e}{m}}\,\frac{V^{3/2}}{r_a} \qquad (2)$$

where h is the height of the cylinders, V is the voltage between the cylinders, and β is a dimensionless function of r_a/r_c tabulated by Langmuir.

The current I_c which flows when the voltage just exceeds the cutoff voltage given in Eq. (1) is shown by Allis[3] to be:

$$\frac{I_c}{h} = \frac{\pi\epsilon_o a_1}{4}\left(\frac{e}{m}\right)^2 r_c^2\left(\frac{r_a}{r_c} - 1\right) B^3 \qquad (3)$$

where a_1 is a function of r_a/r_c (Fig. 9) approximately equal to one.

Eqs. (1), (2) and (3) apply to the dc magnetron.

One equation is of paramount importance for the oscillating magnetron — the Hartree relation. Hartree[4] has shown that there is a critical relation between B and V which establishes a value of V below which oscillation cannot be expected to start. If V is lower than the value specified by Eq. (4), no electrons can reach the anode with an infinitesimal amplitude of rf voltage on the anode; a finite amplitude is required and there is no way for this amplitude to be generated.

$$V = \pi c\,\frac{Br_a^2}{|k|\lambda}\left[1 - \left(\frac{r_c}{r_a}\right)^2\right] - \frac{m}{2e}\,r_a^2\left(\frac{2\pi c}{k\lambda}\right)^2 \qquad (4)$$

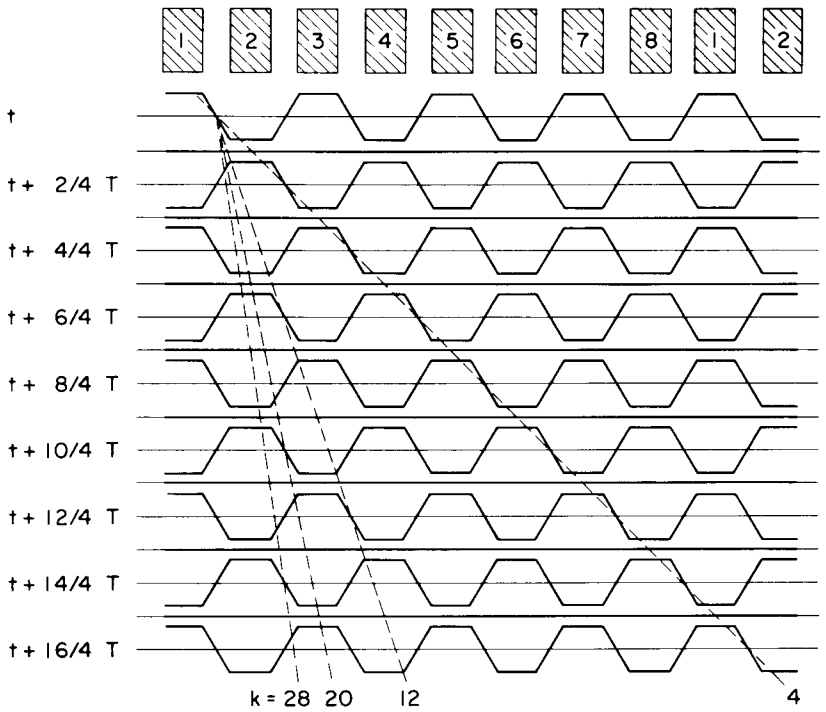


Figure 8. Plot Showing π-Mode Anode Potential Wave at Several Instants in an 8-Resonator Magnetron

Note that this relation depends on $|k|\lambda$, the product of the Hartree mode number and the wave length. The value of anode voltage V given by Eq. (4) is called the Hartree voltage.

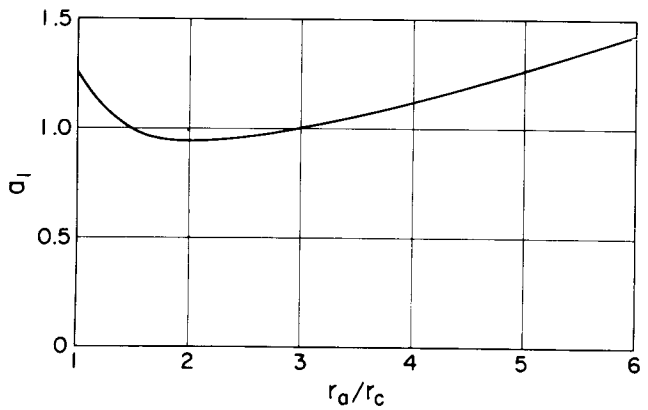


Figure 9. Plot of a_1 as a Function of r_a/r_c

Fig. 10 shows the current vs. voltage characteristic of a magnetron for a fixed magnetic field. The solid line represents the ideal dc case; the current is zero until the Hull cutoff voltage, Eq. (1), is reached. The current then jumps to the Allis current, Eq. (3). As the voltage is increased above the cutoff voltage, the current increases and approaches the Langmuir diode current, Eq. (2). The dashed curve shows the practical dc case. The current begins to flow below cutoff. No satisfactory explanation of the "leakage" current has been made. Initial velocities of electrons and tube asymmetries are not sufficient to explain the effect. The long-dash line shows the current for an oscillating tube. When the voltage reaches the Hartree voltage, Eq. (4), oscillation begins and the current increases above the dc leakage current. It continues to increase above the dc value until the voltage nearby reaches the cutoff value, at which point oscillation ceases. The value of current at this point is empirically given as one-half the Allis current. This value is not very accurate, however, as most tubes cease oscillating at current values considerably below that of half of the Allis current.

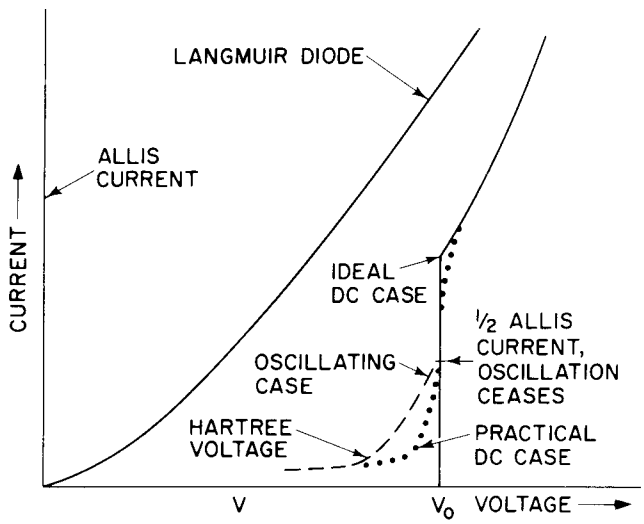


Figure 10. Current-vs.-Voltage Characteristic of a Magnetron for a Fixed Magnetic Field

The Hartree voltage for the various magnetron modes is often plotted against magnetic field. Such a Hartree diagram for an 8-resonator magnetron is given in Fig. 11. This Hartree diagram shows the voltages at which oscillation in the different magnetron modes can start. It is most useful in determining the mode in which the magnetron is operating. The line representing the voltage for a particular mode is termed the Hartree line for that mode. The Hartree mode lines k = n = 4, 3, and 2 are tangent to the Hull cutoff curve. The line for n = 4 gives the voltage at which oscillation starts for the π mode. As the current is increased the voltage rises above the line for n = 4 as indicated by the two lines shown in dashes. As the magnetic field is decreased, the electronic efficiency decreases until it reaches zero at the intersection of the line for n = 4, as indicated by the two lines shown in dashes. As the magnetic field is decreased, the electronic efficiency decreases until it reaches zero at the intersection of the line for n = 4 and the Hull cutoff curve. The value of B and V at this intersection, $\mathcal{B}$, the characteristic magnetic field, and $\mathcal{V}$, the characteristic voltage, are of importance in scaling. Fig. 12 shows a three-dimensional plot of current and voltage vs. magnetic field, combining the information Figs. 10 and 11.

Interaction-Space Design Equations

There are two basic ways to design the interaction space of a magnetron: (1) calculate the design from the basic design equations, or (2) derive a new design from an existing magnetron design in one of several different ways. If a design is available which does not differ too greatly from the desired design, the second method is usually preferable. However, it is often worthwhile to check the design by applying the basic design equations.

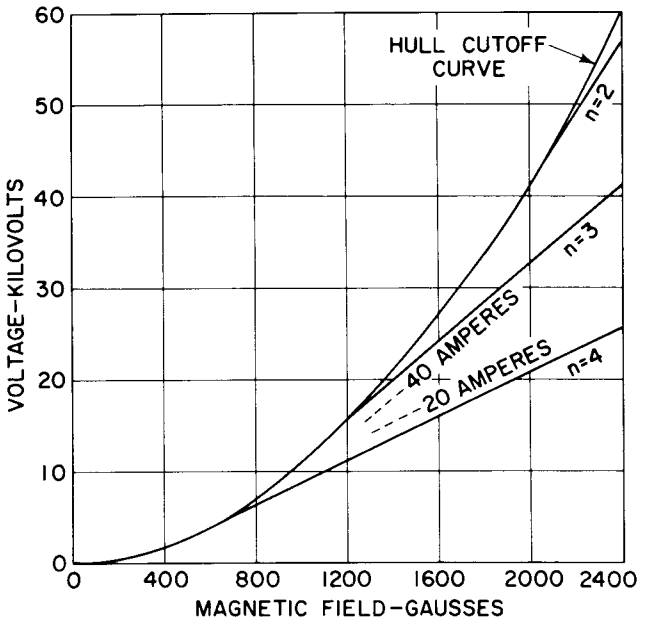


Figure 11. Hartree Diagram for an 8-Resonator Magnetron

First, the basic equations involved in method 1 will be presented; several ways of applying method 2 will be covered later.

The use of the characteristics scale factors,* $\mathcal{B}$, $\mathcal{V}$, $\mathcal{I}$, $\mathcal{P}$, and $\mathcal{G}$, is valuable in both methods. In general, these factors depend upon the Hartree mode number k. For each mode, there are different values of the scale factors. In the design of magnetrons, the π mode values k = n = N/2 are usually desired; therefore, in the equation n will be used instead of |K|.

The characteristic magnetic field $\mathcal{B}$ is the magnetic field at the intersection of the Hartree line and the cutoff curve (see Fig. 11) and is given by:

$$\mathcal{B} = 2\,\frac{m}{e}\,\frac{2\pi c}{n\lambda}\,\frac{1}{1-\left(\frac{r_c}{r_a}\right)^2} \tag{5}$$

or, if $\mathcal{B}$ is in gausses and λ in centimeters

$$B = \frac{21200}{n\lambda\left[1-\left(\frac{r_c}{r_a}\right)^2\right]} \text{ gausses} \tag{5a}$$

Similarly, the characteristic voltage $\mathcal{V}$ is the voltage at the intersection of the Hartree mode line and the cutoff curve:

$$\mathcal{V} = \frac{1}{2}\,\frac{m}{e}\left[\frac{2\pi c}{n\lambda}\right]^2 r_a^2 \tag{6}$$

or

$$\mathcal{V} = 253,000\left(\frac{2\pi r_a}{n\lambda}\right)^2 \tag{6a}$$

The magnetron efficiency is zero at the point $\mathcal{V}$, $\mathcal{B}$ and increases as the operating point moves out along the mode line, i.e., as V and B are increased. The characteristic current $\mathcal{I}$ is the current flowing when B equals $\mathcal{B}$ and V is just greater than $\mathcal{V}$, with no rf voltage present.

$$\mathcal{I} = \frac{2\pi a_1}{\left[1-\left(\frac{r_c}{r_a}\right)^2\right]\left[\frac{r_a}{r_c}+1\right]}\,\frac{m}{e}\left(\frac{2\pi c}{n\lambda}\right)^3 r_a^2\,\epsilon_o h \tag{7}$$

or

$$\mathcal{I} = \frac{8440\ a_1}{\left[1-\left(\frac{r_c}{r_a}\right)^2\right]\left[\frac{r_a}{r_c}+1\right]}\left(\frac{2\pi r_a}{n\lambda}\right)^3 \frac{h}{r_a} \text{ amperes} \tag{7a}$$

where a_1 is a function of r_a/r_c (see Fig. 9)

$\mathcal{P}$ is the characteristic power:

$$\mathcal{P} = \mathcal{I}\mathcal{V} \tag{8}$$

The characteristic conductance $\mathcal{G}$ is

$$\mathcal{G} = \mathcal{I}/\mathcal{V} \tag{9}$$

There are a number of ways of applying these equations to the design of the interaction space. The method given here starts with the specification of the dc magnetic field B. (Alternative methods will be found in

*Page 416 of reference 5.

section 10 of Ref. 5) The method is as follows:

Step 1: Specify values of λ, V, and P_o. Assume a value of B which can reasonably be obtained. Estimates of the anode height and radius will have to be made to judge what value is reasonable.

Step 2: Assume a value of N. Calculate r_a/r_c from

$$\sigma = \frac{r_a}{r_c} = \frac{N - 4}{N + 4} \tag{10}$$

Step 3: Calculate $\mathcal{B}$ from Eq. (5a).

Step 4: Calculate $\mathcal{V}$ from Eq. (4), Hartree relation, which can be rewritten in terms of $\mathcal{V}$ and $\mathcal{B}$ as

$$\frac{V}{\mathcal{V}} = 2\,\frac{B}{\mathcal{B}} - 1 \tag{4a}$$

Step 5: Calculate r_a from Eq. (6a). Check to see that $2r_a/\lambda < 0.3$ for a strapped tube or < 0.6 for the rising-sun type.
(This inequality must be satisfied to insure adequate mode separation.)

Step 6: Check the value of electronic efficiency by means of Slater's equation:

$$\eta_e = 1 - \left(\frac{\mathcal{B}}{B}\right)^2 \frac{V}{\mathcal{V}} \tag{11}$$

If η_e is low, a larger value of N or a higher value of B must be used. (Because this equation is based on very simple assumptions, it may give inaccurate results. This is a disadvantage of this method compared to scaling from a known tube.)

Step 7: Assume a reasonable value for the circuit efficiency η_c and calculate I from

$$I = \frac{P_o}{V \eta_e \eta_c} \tag{12}$$

Step 8: Assume a value of anode height h, keeping in mind the problem of attaining the required magnetic field and the restriction that $h/\lambda < 0.3$ for a strapped tube or < 0.6 for a rising-sun type (to insure adequate mode separation). Calculate the cathode current density Jc from

$$J_c = \frac{I}{2\pi r_c h} \tag{13}$$

If J_c is higher than the proposed cathode can provide, a larger value of h or of N must be used.

Step 9: Select a value of μ (the relative slot width) in the range $0.5 > \mu > 0.3$

$$\mu = W \Big/ \frac{2\pi r_a}{N} \tag{14}$$

Satisfying this condition insures a good rf field pattern in the interaction space. Fac-

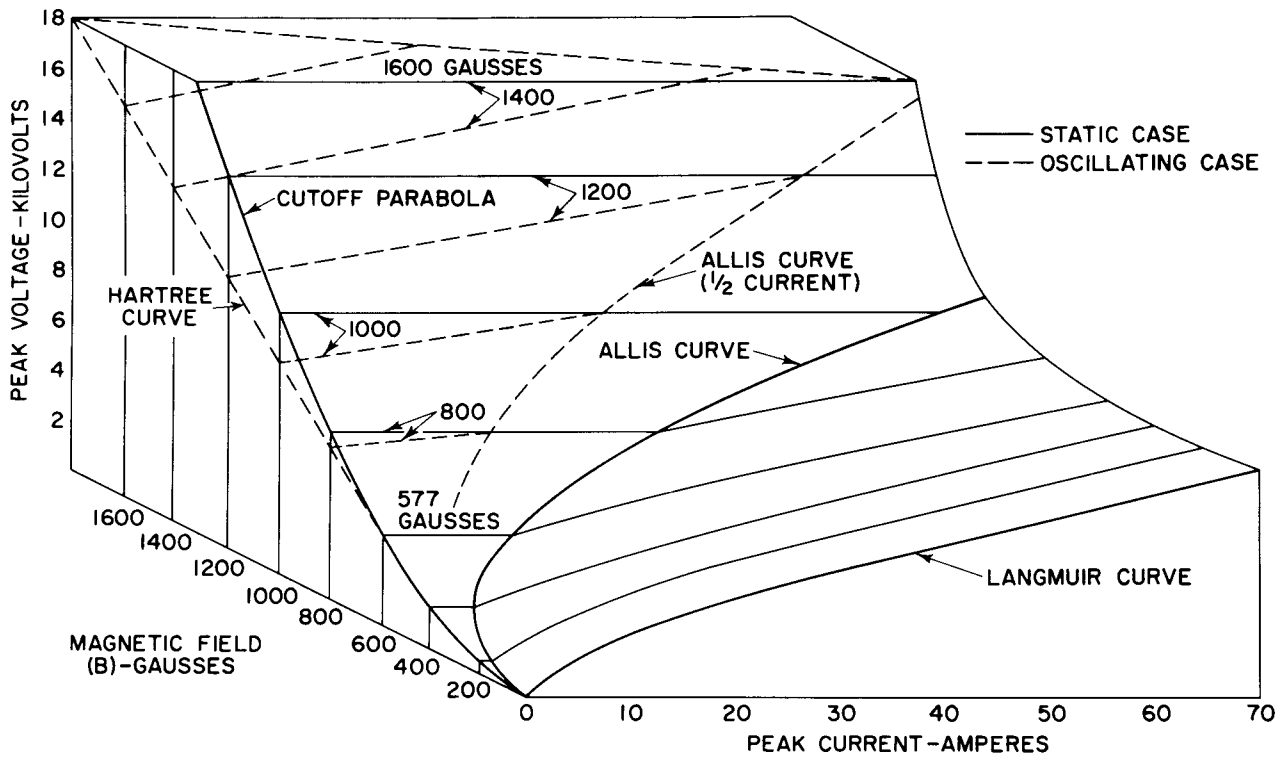


Figure 12. Three-Dimensional Plot of Current vs. Voltage vs. Magnetic Field for a Magnetron

tors to be considered in choosing μ are thermal dissipation, mechanical strength, and its effect on the resonant circuit and on the slot conductance.

Two factors remain undetermined, the slot conductance G_1 and the ratio of load conductance to maximum power load conductance g/g_0. The slot conductance is a characteristic of the resonant circuit. Methods of calculating G_1 are given in chapter 4 of Ref. 5. However, the difficulty of accurately calculating G_1 is such that the usual practice is to ignore it in the design of magnetrons. Fortunately, the operation is not critical with respect to G_1; for a particular design, the electronic efficiency varies from 50 to 60 per cent and back down to 50 per cent, as G_1 is varied over a 5 to 1 range. The ratio g/g_0 indicates at what fraction of the load for maximum power the magnetron operates and, therefore, is part of the output circuit problem.

The resonator impedance (which is a function of G_1 and the other resonator characteristics) and the value of g/g_0 are important in magnetrons designed for very high power output. It is also important that the rf voltage be kept as low as possible in the design of very high power tubes, in order that the electrons may interact efficiently and, in addition, so that arcing between the vane tips and between the straps is prevented. The impedance appearing at the vane tips should, therefore, be made as low as possible.

Magnetron Scaling

Given the interaction-space design of a magnetron of known characteristics, the design of a magnetron of different characteristics but with the same electronic efficiency can be obtained by using scaling equations. This method was originated before the design equations discussed earlier were developed to their present state and is still used because of its convenience. It should be emphasized that the application of the scaling equations gives a design for a new magnetron of the same electronic efficiency as the original magnetron. Thus, if a magnetron for long wave length is scaled to provide a shorter-wave length magnetron, the over-all efficiency will usually decrease because the resonant circuit for the short wave length will usually have a lower circuit efficiency than that for the long wave length.

The scaling equations given below derive directly from the equations for the characteristic scale factors given previously. This derivation is equivalent to requiring that the ratio of the time required for an electron to move between two points in the original magnetron to the time required for an electron to move between corresponding points in the new magnetron should be proportional to the ratio of the wave lengths. It also requires that the electron execute the same number of convolutions in moving between the two points in each case. This requirement insures that the electronic efficiency remains constant. In earlier scaling equations,[6] this requirement was not included and the electronic efficiency, therefore, could change considerably if a large change in N were made. If N is the same for both designs, the equations reduce to the original scaling equations derived by Clogston[7] as shown in the forthcoming paragraphs.

Eq. (10) for the ratio of cathode radius to anode radius is often ignored so that the cathode can be made larger to obtain improved mode stability (at the cost of reduced electronic efficiency). For this reason, the equations are given in terms of r_c and r_a instead of N, so that they may easily be applied even if Eq. (10) is not satisfied.

Define the following ratios where the subscript 1 refers to the original magnetron and the subscript 2 refers to the new magnetron.

$$\alpha = \frac{r_{a2}}{r_{a1}} \qquad \theta = \frac{N_2}{N_1}$$

$$M = \frac{\lambda 2}{\lambda 1} \qquad \phi = \frac{h_2}{h_1}$$

$$\delta = \left(\frac{N_2 - 4}{N_2 + 4}\right)\left(\frac{N_1 + 4}{N_1 - 4}\right)$$

if Eq. (10) is satisfied, the cathode radius is given by:

$$r_{c2} = \alpha\, \delta\, r_{c1} \tag{15}$$

If Eq. (10) is ignored, r_{c2} may be arbitrarily chosen; its value may be as large, perhaps, as 10 per cent higher than $r_{a2}\,(N_2 - 4)/(N_2 + 4)$.

The operating voltage of the new tube is

$$V_2 = V_1\left(\frac{\alpha^2}{\theta^2\, M^2}\right) \tag{16}$$

and the magnetic field of the new tube is given by

$$B_2 = B_1\left(\frac{1}{\theta\, M}\right)\frac{1-\left(\frac{r_{c1}}{r_{a1}}\right)^2}{1-\left(\frac{r_{c2}}{r_{a2}}\right)^2} \tag{17}$$

The current of the new tube corresponding to current I_1 of the original tube is

$$I_2 = I_1\left(\frac{a_{1,2}}{a_{1,1}}\right)\left(\frac{\alpha^2\phi}{\theta^3\, M^3}\right)\frac{\left[1-\left(\frac{r_{c1}}{r_{a1}}\right)^2\right]^2}{\left[1-\left(\frac{r_{c2}}{r_{a2}}\right)^2\right]^2}\left\{\frac{1+\frac{r_{a1}}{r_{c1}}}{1+\frac{r_{a2}}{r_{c2}}}\right\} \tag{18}$$

and the corresponding cathode current density is

$$J_{c2} = J_{c1}\left(\frac{a_{1,2}}{a_{1,1}}\right)\left(\frac{\alpha^3}{\theta^3 M^3}\right)\frac{\left[1-\left(\frac{r_{c1}}{r_{a1}}\right)^2\right]^2}{\left[1-\left(\frac{r_{c2}}{r_{a2}}\right)^2\right]^2}\left(\frac{r_{c1}+r_{a1}}{r_{c2}+r_{a2}}\right) \quad (19)$$

The generated power P_2 of the new tube corresponding to the generated power P_1 of the original tube is given by

$$P_2 = P_1\left(\frac{a_{1,2}}{a_{1,1}}\right)\left(\frac{\alpha^4\,\phi}{\theta^5 M^5}\right)\frac{\left[1-\left(\frac{r_{c1}}{r_{a1}}\right)^2\right]^2}{\left[1-\left(\frac{r_{c2}}{r_{a2}}\right)^2\right]^2}\left\{\frac{1+\frac{r_{a1}}{r_{c1}}}{1+\frac{r_{a2}}{r_{c2}}}\right\} \quad (20)$$

and the power output is

$$P_{o2} = P_2\,\eta_e\,\eta_{c2} \quad (21)$$

If $N_2 = N_1$, the equations are much simpler and are identical to those derived by Clogston,[7] i. e.,

$$V_2 = V_1\left(\frac{\alpha}{M}\right)^2 \qquad r_{c2} = r_{c1}\,\alpha$$

$$B_2 = B_1/M \qquad J_{c2} = J_{c1}\frac{\alpha}{M^3} \quad (22)$$

$$I_2 = I_1\frac{\alpha^2\phi}{M^3} \qquad P_2 = P_1\frac{\alpha^4\phi}{M^5}$$

The application of the equations is straightforward although it involves a cut-and-try procedure to obtain the desired characteristic of the new magnetron.

Reduced-Performance Charts; Generalized Hartree Diagrams

A plot of contours of magnetic field, power output, efficiency, and sometimes frequency, in the voltage-current plane for a particular magnetron is known as its performance chart. Fig. 13 shows a performance chart of a typical pulse magnetron. The performance chart applies to a particular tube geometry (specified by N, μ, r_c, and r_a) operating with a particular loading, i. e., a specific resonant circuit and output circuit connected to a specific load, usually a matched load.

The preceding explanation has shown that from a given magnetron, operating at a particular operating point on the performance chart, the application of the scaling equations will give the characteristics of a new magnetron operating at a corresponding operating point. If a number of operating points on the performance chart of the original magnetron are all scaled to a new magnetron of a particular geometry and loading, the performance chart of the new magnetron may be predicted. It should be noted that μ is not involved in the design equations or the scaling equations. Therefore, if the results are to be absolutely accurate, μ must be the same in both the original and the new magnetrons.

Calculation of the new performance chart can be simplified by plotting the universal or reduced-performance chart of the original tube. Fig. 14 shows the reduced performance chart of the type 2J32 magnetron whose conventional performance chart was given in Fig. 13. In Fig. 14, the reduced magnetic field, power output, and efficiency are plotted on the reduced voltage-current plane. The reduced or dimensionless variables are given by

$$b = \frac{B}{\mathcal{B}} \qquad i = \frac{I}{\mathcal{I}} \qquad (23)$$

$$v = \frac{V}{\mathcal{V}} \qquad p = \frac{P}{\mathcal{P}}$$

where B, V, I, and P are the magnetic field, voltage, current, and power generated, measured experimentally, while $\mathcal{B}$, $\mathcal{V}$, $\mathcal{I}$, and $\mathcal{P}$ are the characteristic magnetic field, voltage, current, and power calculated from Eqs. (5, 6, 7, and 8).

Having derived this reduced-performance chart, the performance chart of any new magnetron may be obtained by calculating the characteristic factors $\mathcal{B}$, $\mathcal{V}$, $\mathcal{I}$, and $\mathcal{P}$ and reapplying the above equations to obtain B, V, I, and P_e for the new tube. The electronic efficiency remains unchanged, provided the original and the new tube have the same μ and loading. Therefore, the power output of the new tube is $P\,\eta_c$.

This reduced-performance chart represents all magnetrons having the same μ and loading as the original tube for which it was plotted. That is to say, that if the effect of changes in μ or in the loading are ignored, then one reduced-performance chart represents all magnetrons of the traveling-wave type. The performance chart for any new magnetron may be obtained from it by calculating the characteristic factors, multiplying them by the reduced variables, and taking into account the circuit efficiency in calculating the power output and over-all efficiency.

A Universal or Generalized Hartree Diagram is shown in Fig. 15; this diagram is useful in determining whether a tube is operating in the π mode. The diagram is plotted on a reduced-voltage-vs.-reduced-magnetic field plane so that it may be applied to tubes of any wave length and anode geometry. The operating point of a tube should lie just above the mode-number line k of the mode in which the tube is operating. Thus, for a tube operating in the π mode, the operating point should lie just above the line k = N/2.

The Generalized Hartree Diagram may also be used for designing a magnetron by a cut-and-try method. If values of N, λ, B, r_c, and r_a are assumed, the normalized variables may be calculated and the operating voltage obtained from the chart by the use of the proper mode line, k = N/2. Operation at points near the Hull cutoff curve must be avoided because the efficiency in this region will be very low. For good electronic efficiency, operating points to the right of the dotted curve should be selected.

MAGNETRON RESONANT CIRCUITS

Types of Resonant Circuits

For the magnetron to generate rf power, there must be an rf field with which the electrons in the space charge can interact. This field must be of the desired frequency and of the proper configuration for efficient generation of rf power. The resonant circuit provides this field by storing a quantity of energy, obtained over a number of rf cycles. The resonant circuit must be a stable frequency-determining element and must provide fields extending into the interaction space, where the electrons can interact with them.

The resonant circuit of a magnetron oscillator may be thought of as a filter circuit with a narrow pass band which permits only the desired frequency component in the electron stream to build up. As the magnetron starts oscillation, the resonant circuit absorbs and stores energy at this frequency. This storage results in increased rf electric field with which the electrons interact. The rf field, in turn, causes the electrons to bunch in such a way that their energy in the desired frequency component is increased. Thus, the resonant system can absorb more energy, present still higher fields to the electrons, etc. The start of oscillations is, as in the case of other oscillators, due to random fluctuations.

There are two major types of resonant circuits used in traveling-wave magnetrons. The strapped system, shown in Fig. 16, consists of a number of cavities having the same resonant frequency and having the front of alternate segments connected by straps to increase the frequency separation of the desired π mode from the other modes. The purpose of using many cavities is to reduce the required magnetic field and voltage for a given frequency and anode diameter. The cavities effectively operate in parallel and the frequency of the system is approximately that of a single cavity. The equivalent circuit of the strapped anode is shown in Fig. 17.

The second major type of resonant circuit for traveling-wave magnetrons is the rising-sun type shown in Fig. 18. The rising-sun system consists of N resonant cavities of two different frequencies. The large and small cavities alternate around the anode; their frequencies are so chosen to give optimum mode separation without the necessity of strapping. The rising-sun system has the advantage of permitting twice as large an anode diameter and anode length as the strapped anode, before other modes become troublesome. It is therefore used for very high-frequency and/or high-power tubes.

Because the strapped resonator system is the more common type, a method of design for this type follows.

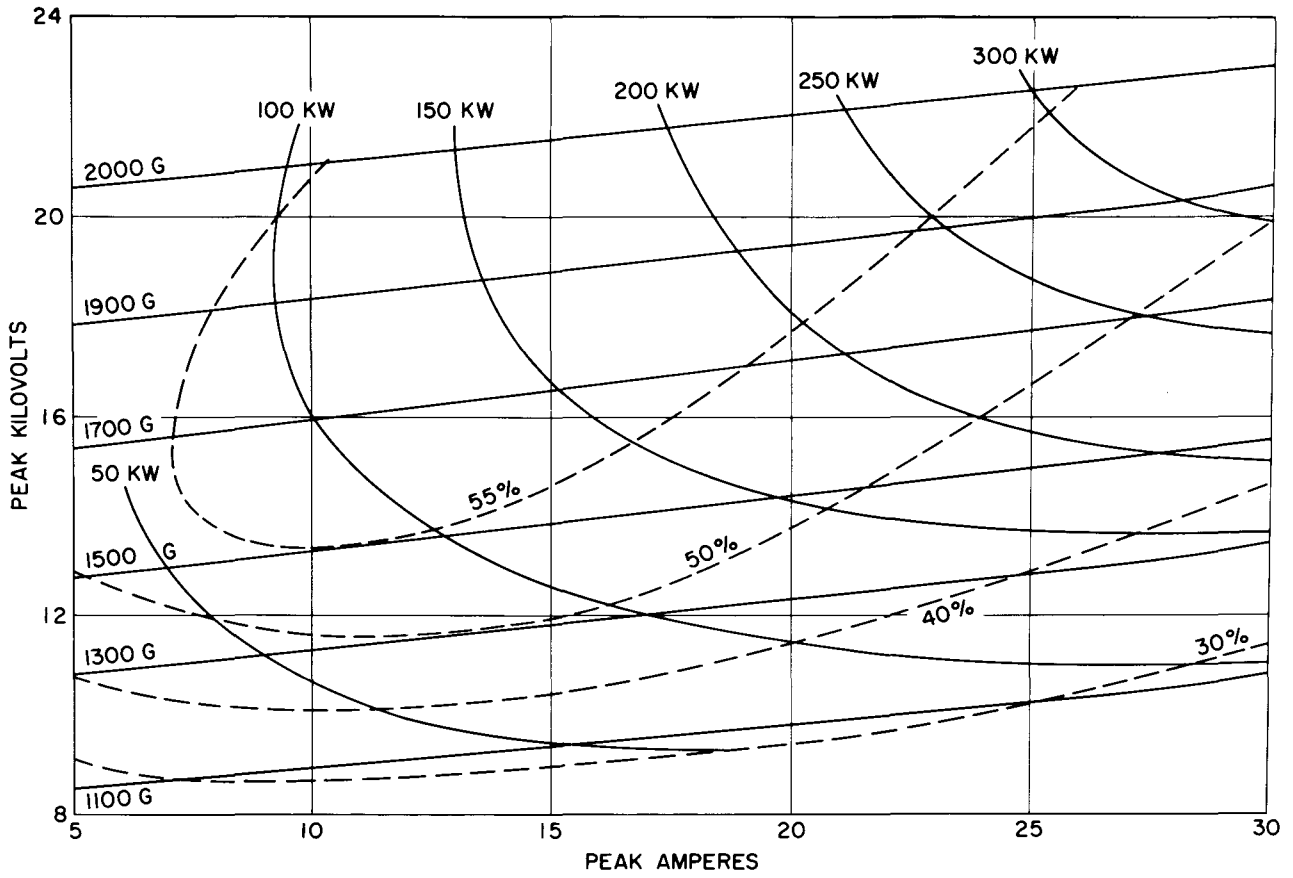


Figure 13. Performance Chart of a Typical Pulse Magnetron (Type 2J32) (frequency, 2800 Mc; pulse duration, 1 μsec; prr, 1000 pps; pulling figure, 12 Mc)

Design information on rising-sun systems may be found on page 470 of Ref. 5.

Design of Strapped-Resonator Systems

The basic requirement to be satisfied in designing the resonator system is to obtain values compatible with the interaction space design for the following four characteristics: μ, the relative slot width; G_1, the slot conductance; Q_0, the unloaded Q, and λ, the resonant wave length. The handling of μ and G_1 was discussed earlier. Q_0 should, of course, be as high as possible to give the maximum circuit efficiency (for a given Q_{ext} or output loading). Q_0 increases with wave length and varies inversely as the number of resonators. It also varies inversely as the relative slot width and the slot conductance. In general, therefore, there is but a limited control over the value of Q_0 because the range of these parameters is limited by other considerations. Approximate formulas for the calculation of Q_0 for strapped anodes is given on page 466 of Ref. 5.

The most important resonator characteristic, the resonant wave length, is now to be determined. The basic method for determining the wave length is to solve Maxwell's equations in the resonator region. Because the boundary conditions at the junctions with the interaction space and the end space regions are complicated and require approximations, the effort of solving Maxwells's equations usually is not justified. Instead, it is normal practice to use approximate formulas to determine the wave length. The results are then verified by constructing a model anode whose frequency can be measured by cold test methods (See Chapter 18, Ref. 2). The approximation which follows is due to W. V. Smith (Ref. 3) and is accurate to within 5 per cent.

The equivalent circuit of a magnetron operating in the π mode can be represented by a simple parallel circuit of inductance L and capacitance C, whose values may be calculated from the tube geometry. If the magnetron is strapped, an additional shunt capacitance C_s is added which may be calculated from the strap geometry. The wave length λ is then obtained from:

$$\lambda = \sqrt{0.0355\ L\ (C + C_s)} \qquad (24)$$

where L is in micromicrohenries and C in micromicromicrofarads.

Values of L and C may be obtained by constructing a scale drawing of the cavity, as shown in Fig. 19. The area of the cavity is divided into two areas: A_2, the area from the vane tips to a distance s/3, and A_1, the remaining area. Angles ϕ_1 and ϕ_2 are also constructed and measured. C and L are then obtained from:

$$C = 0.090\ N\ (0.8 + 0.7\ \text{Cot}\ \phi_2 + 0.2\ \text{Cot}\ \phi_1 + 0.02\ s/w)\ (h + 0.3\ w)\ \text{micromicrofarads} \qquad (25a)$$

$$L = \left(\frac{12{,}600}{N}\right)\left(\frac{A_1 + 0.3\ A_2}{h + s/3}\right) \text{micromicrohenries} \qquad (25b)$$

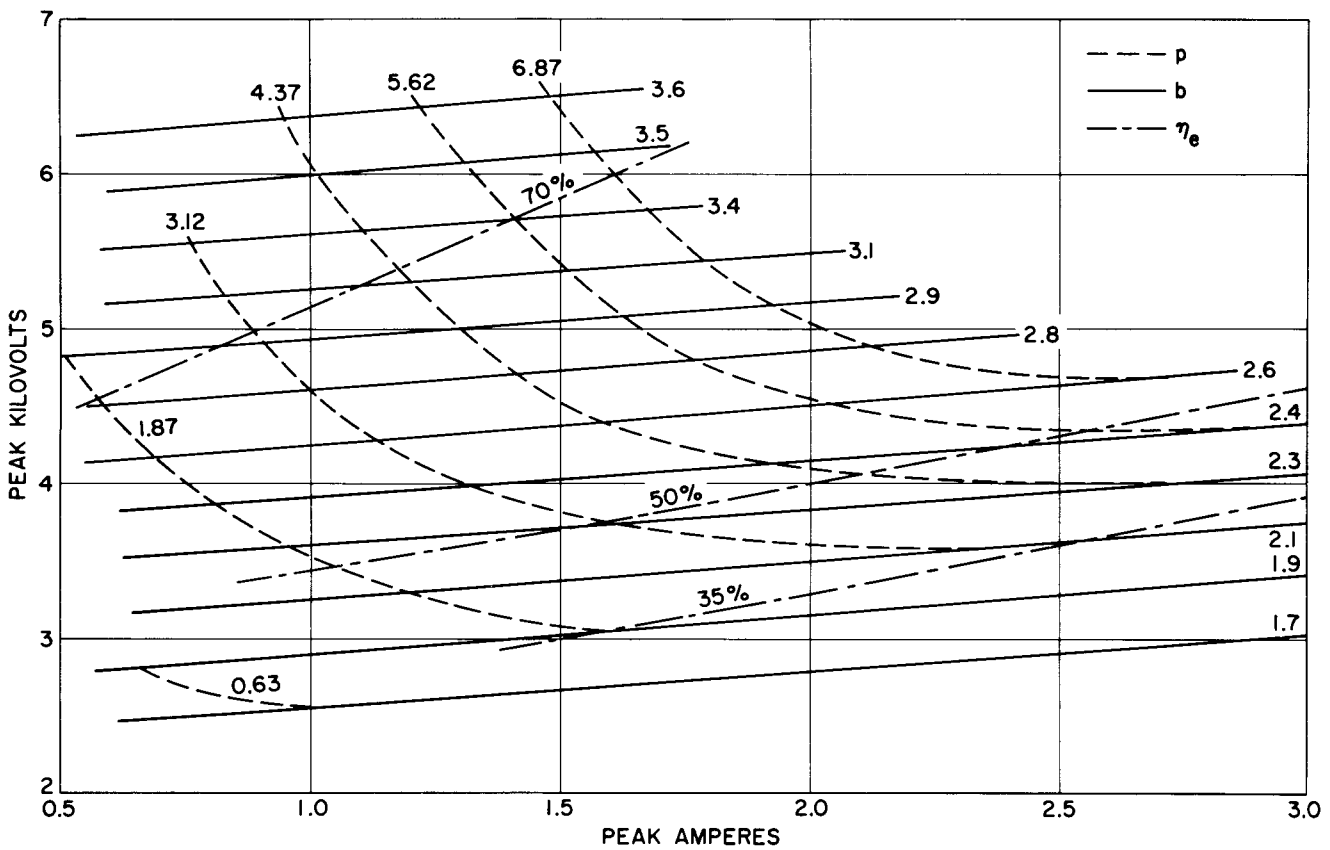


Figure 14. Reduced Performance Chart of Type 2J32 Magnetron (N = 8; σ = 0.38; μ = 0.37; g = 0.51)

where h is the axial height of the resonator.

The strap capacitance is calculated by using standard capacitance formulas. For double-ring strapping, the capacitance value will include the capacitance between the two straps (multiplied by two if there is a set of straps at each end of the anode) and the capacitance between each strap and the vanes which are not connected to the strap. Formulas for standard resonator geometries (for calculating C and L), as well as formulas for some of the more common strap geometries, are given in Ref. 8.

MAGNETRON TUNING

Single Tuning Cavity

The most direct method of mechanically tuning[9] the frequency of a magnetron is to change the inductance or capacitance of its anode resonators. Fig. 20A illustrates two conventional methods used for this purpose. The "crown-of-thorns" method utilizes a number of pins that can be inserted into the back part of the magnetron resonators; these pins decrease the inductance of the resonant system and thereby tune the tube to a higher frequency. Magnetrons of this type are capable of producing a plate power output of approximately 500 kilowatts at a duty factor of 0.001 and can be tuned over a frequency range of 10 per cent within the X-band. Figure 20B illustrates the technique used to change the effective capacitance of the anode resonator. The bottom end of the "cookie-cutter" tuner is made to penetrate between the strap rings of the magnetron. By this method, the effective strap spacing is decreased; the accompanying increase in capacitance results in a lower resonant frequency. Although these methods can provide satisfactory tuning ranges, other performance characteristics, such as stability, may be adversely affected.

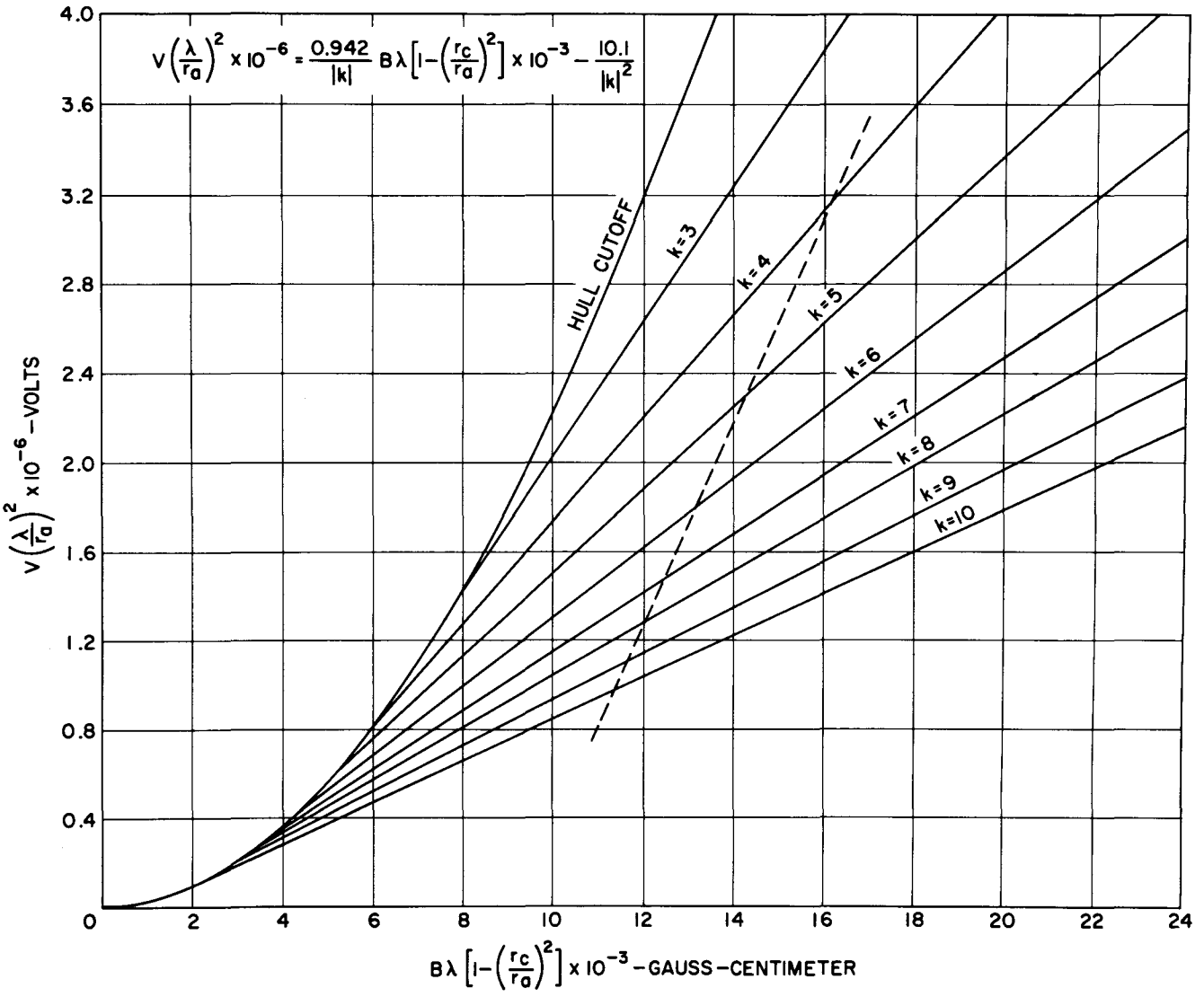


Figure 15. Universal (Generalized) Hartree Diagram

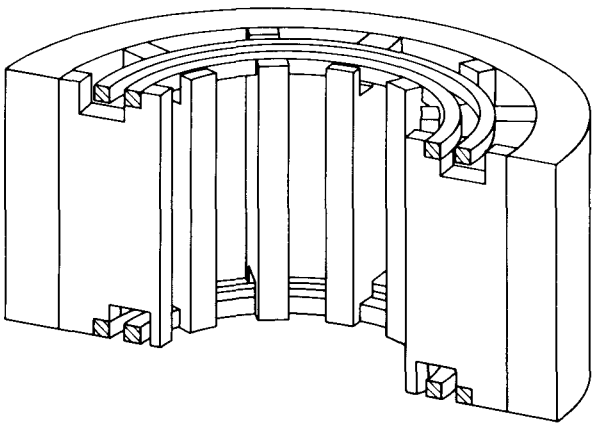
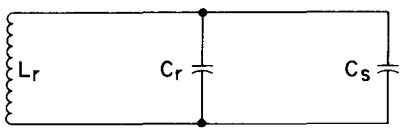


Figure 16. Double-Ring Strapped-Anode Block (Section)



L_r = RESONATOR INDUCTANCE
C_r = RESONATOR CAPACITANCE + FRINGING CAPACITANCE
C_s = STRAP CAPACITANCE

Figure 17. Equivalent Circuit of Strapped Anode

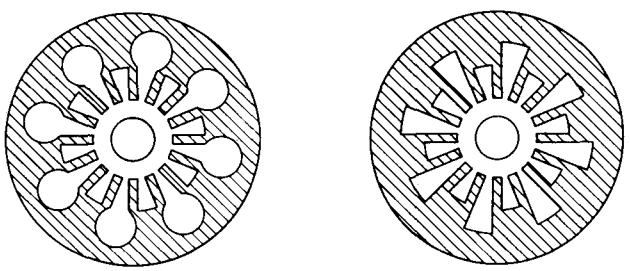


Figure 18. Rising-Sun Types of Anodes

Another method of changing the frequency of a magnetron is to vary the reactance coupled into the anode circuit by means of one or more external tuning cavities connected to anode resonators. This method is well suited for high-power tubes because it does not restrict anode cooling or require small dimensions which can lead to voltage breakdown problems. A commercial, high-power, X-band magnetron tuned in this manner has demonstrated nearly constant power output over a wide tuning range with simultaneous attainment of low thermal drift, good spectrum shape, and excellent mode stability even at high rates of voltage rise.

Tuning ranges slightly greater than the frequency separation between the π mode and the (N/2)-1 mode have been obtained at high peak power output by use of four tuning cavities symmetrically arranged with respect to the anode, i.e., 90 degrees between adjacent tuning cavities. The use of four cavities reduces field pattern distortion and improves mode stability. High-power tubes tuned with a single external cavity seem to be limited to bandwidths of the order of one-fourth the mode separation between the π mode and the (N/2)-1 mode. This approximate limit is set by a decrease in electronic efficiency, increased heating of the cathode by back-bombardment, and moding — mostly as a result of field pattern distortion. However, certain continuous-wave (CW) magnetrons operating at a relatively low current level and power output, successfully use single-cavity tuning to cover a frequency range almost as great as the frequency separation between the π mode and (N/2)-1 mode. Performance charts for single-cavity, high-power, pulsed tubes also confirm that much wider tuning can be obtained with good electronic efficiency when the power output is relatively low.

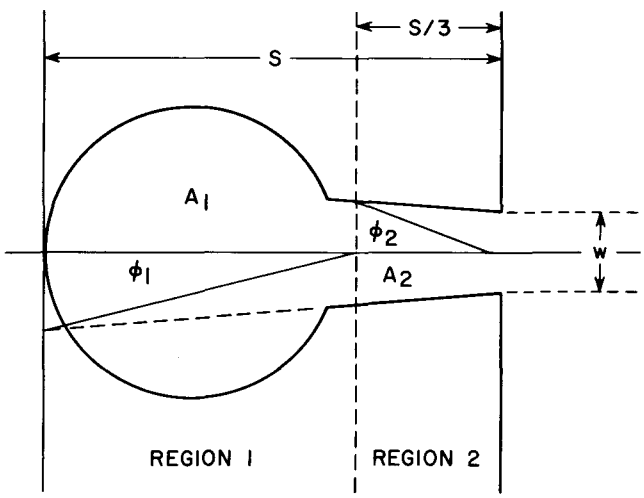


Figure 19. Resonator Geometry as Drawn for Calculation of Frequency

An important feature of cavity-tuned magnetrons is their relatively small thermal frequency drift when the tuner position or the input power is changed. These magnetrons usually have thermal characteristics superior to magnetrons which are tuned by elements having small clearance to the anode, because cavity tuners dissipate little energy and the consequent heating causes little movement of the tuner parts. Because the field configuration in the cavity is well understood, it is relatively easy to avoid extraneous tuner resonances and the associated problems of heating and drop in output power.

Description. Although the use of a single coupled cavity[9, 10, 11, 12] has given large tuning ranges only at relatively low power, a study of its characteristics is important because it helps in understanding some of the general considerations applicable to the design of multiple-cavity tubes.

Consider the magnetron anode and tuning arrangement shown in Fig. 20C. The double-strapped vane-type magnetron is coupled to a ridged wave guide tuning cavity through a slot in the resonator diametrically opposite to the output transformer. This cavity has the

desirable features of low characteristic impedance, wide band-width free from higher mode interference, and an adequately high Q. It can be constructed with the width of the ridge equal to the anode height, thereby permitting the use of a thin coupling iris which is relatively broadband. Because most of the energy in the cavity is in the region between the ridges, a movable short circuit between them provides a simple and effective tuner.

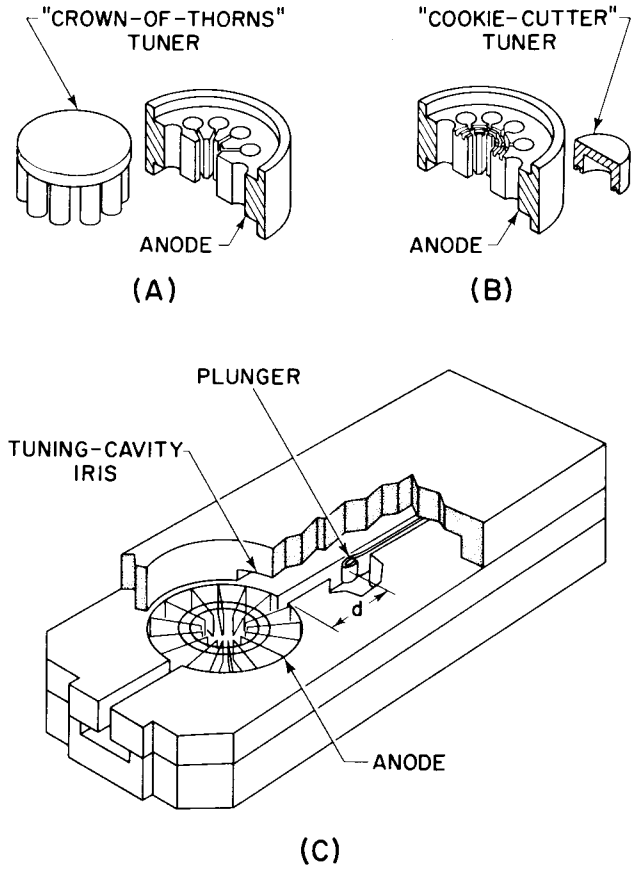


Figure 20. Magnetron Tuning: (A) Crown-of-thorns inductive tuning element and anode; (B) Cookie-cutter capacitive tuning element and strapped anode; (C) Single-cavity tunable magnetron

Equivalent Circuit. Microwave devices such as tunable magnetrons are usually large compared to the free-space wave length at which they operate and therefore involve parameters of a distributed nature. An equivalent circuit with lumped constants and having the same admittance or impedance function as the device can be synthesized according to network theory, but the usefulness of the circuit is restricted unless it has physically realizable parameters that can be directly related to parts of the device. Fortunately, for the case of the single-cavity tunable magnetron, there is a good equivalent circuit which is relatively simple.

An equivalent circuit, neglecting losses, for the single-cavity tunable magnetron is shown in Fig. 21. This circuit is an extension of one used previously by M. A. Herlin[10] and is valid over a much broader frequency range because all anode modes can be taken into account, each mode being represented by one branch of a series inductance-capacitance (LC) circuit. The iris between anode and tuning cavity can be represented by a parallel combination of L and C; if, however, the iris is resonant far from the frequency range of interest, a single shunt-reactance element is a good approximation. The ridge wave guide tuning cavity is represented by a section of wave guide terminated in a movable short circuit and having a characteristic admittance Y_o.

Derivation of an equation relating plunger position d to the resonant frequency of the system, now that the equivalent circuit has been established, is a straightforward network problem using the following relation which applies at resonance.

$$B_m + B_i + B_c = 0 \tag{26}$$

B_m, B_i, and B_c are, respectively, the susceptances of the magnetron anode, coupling iris, and a length d of tuning cavity terminated in a short circuit. These susceptances are given by:

$$B_m = \sum_{m=1}^{N/2} \frac{\sqrt{\frac{C_m}{L_m}}}{\lambda/\lambda_m - \lambda_m/\lambda} \tag{27}$$

$$B_i = \sqrt{\frac{C_i}{L_i}} \left(\lambda_i/\lambda - \lambda/\lambda_i\right) \tag{28}$$

$$B_c = -Y_o \cot\left(\frac{2\pi d}{\lambda_g}\right) \tag{29}$$

$$\lambda_m = 2\pi c \sqrt{L_m C_m} \tag{30}$$

$$\lambda_i = 2\pi c \sqrt{L_i C_i} \tag{31}$$

where

λ = free-space wave length

Y_o = characteristic susceptance of tuning cavity

λ_g = guide wave length in tuning cavity

c = velocity of light

Substituting Eqs. (27), (28), and (29) into Eq. 26 and solving for d gives

$$d = \frac{\lambda_g}{2\pi} \cot^{-1}\left[\sum_{m=1}^{N/2} \frac{\frac{1}{Y_o}\sqrt{\frac{C_m}{L_m}}}{\lambda/\lambda_m - \lambda_m/\lambda} + \frac{1}{Y_o}\sqrt{\frac{C_i}{L_i}}\left(\lambda_i/\lambda - \lambda/\lambda_i\right)\right] + \frac{n\lambda_g}{2} \tag{32}$$

where n = 0, or a positive integer

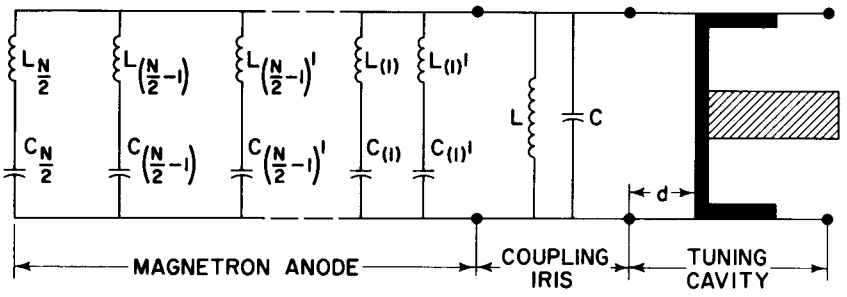


Figure 21. Equivalent Circuit for a Single-Cavity Tunable Magnetron

Using reasonable values for the parameters in Eq. (32) leads to curves of d vs. λ of the type shown in Fig. 22. These curves are commonly referred to as tuning curves.

Tuning Curves. Slater[13] has derived, by the solution of field equations, an expression similar to Eq. (32) for the tuning of a single multiple-resonant cavity tuned by a coupled wave guide with a movable short circuit. Because the magnetron anode can be considered as a multimode cavity, the general observations regarding the tuning curves for the magnetron and the multimode cavity are the same. As in the case of the multimode tunable cavity, a valuable insight into the characteristics of the tunable magnetron can be obtained from these curves.

From Fig. 22 it is seen that, when the plunger is at $d = n\lambda_g/2$, the resonant wave lengths of the circuit are those of the anode alone, the curves being continuous between adjacent anode modes and progressing from one mode to the next as the plunger is moved a distance $\lambda_g/2$. If the field pattern of an anode block is observed as the tube is tuned from the point B to the point C, the pattern will initially be that of the π mode, changing continuously as d is decreased, and finally assuming the pattern of the (N/2)-1 mode. Thus, the field patterns become distorted when a mode is tuned; the amount of distortion is related to the mode separation and amount of tuning. Because field distortion is one of the main limitations on single-cavity-tuned tubes, the widest tuning ranges are obtained by operating about the π-mode resonance and by using a structure having a large mode separation. It also follows that expressing the tuning range as a percentage of the mode separation is meaningful, because it permits a better comparison of the performance of tubes operating at different frequencies and with different mode separations.

It is important to know the distribution of stored energy in the resonant system (as will be discussed later) in order to determine in which of several possible modes of oscillation a tunable magnetron will operate. An estimate of this stored energy can be obtained from the slope of the tuning curves. If most of the energy is in the tuning cavity, the tuning curves will be steep because a perturbation of the boundary (such as a slight motion of the plunger in a part of the system that contains a considerable portion of the energy) will cause a large change in frequency. On the other hand, the nearly horizontal parts of the tuning curves like those shown at the (N/2)-1 and (N/2)-2 modes, indicate that nearly all of the energy is in the anode rather than in the tuning cavity.

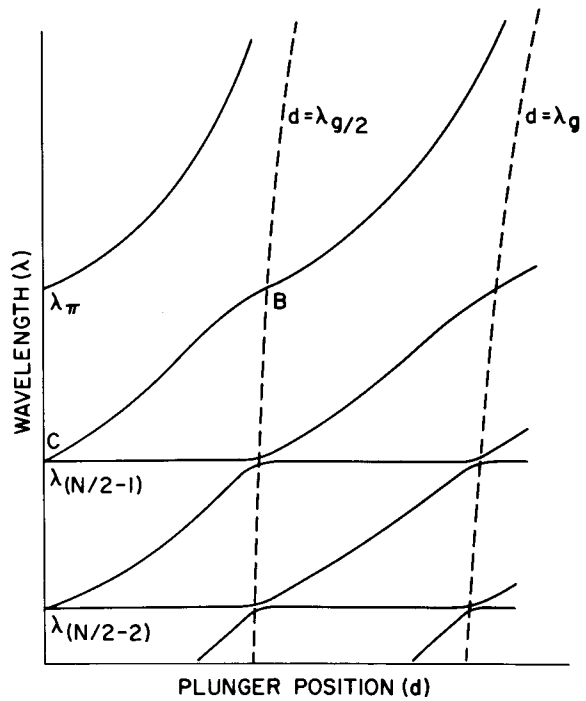


Figure 22. Tuning Curves of a Magnetron with a Single Coupled Tuning Cavity

Another quantity closely related to the distribution of energy is the degree of coupling of the anode to the tuning cavity for different modes. The degree is indicated by the slope of the tuning curves at wave lengths corresponding to anode resonances and by the value of $(1/Y_0)\sqrt{L_m/C_m}$ in Eq. (32). Denoting this term by $1/Q_{ext\ m}$, it is seen that in this form Y_0 appears as the loading on the series resonant circuit formed by L_m and C_m, the inductance and the capacitance of the magnetron anode. Resonances that do not tune will have a very high $Q_{ext\ m}$ indicating, as it should, a very light coupling of the mode to the tuning cavity.

From Eq. (32) it should be noted that, at an anode resonance, the term B_1 is negligible compared to the term B_m. Although the iris does not affect the shape of the tuning curves in this immediate frequency range, it is important at frequencies far removed from an anode

resonance. In these regions, an inductive iris will displace the curves to the right while one that introduces a shunt capacitance will displace the curves to the left.

To obtain precise values of the parameters for the equivalent circuit of a particular tunable magnetron, it is necessary to use information from its experimental tuning curves. From these values, the anode resonant frequencies and the slope of the curves at the resonances can be obtained. Differentiating Eq. (32) with respect to λ and evaluating at the anode resonances gives

$$\left(\frac{d\,d}{d\lambda}\right)_{\lambda=\lambda_m} = \frac{\lambda_{gm}}{\lambda_m}\left[\frac{Y_o}{\pi}\sqrt{\frac{L_m}{C_m}} + \frac{n}{2}\left(\frac{\lambda_{gm}}{\lambda_m}\right)^2\right] \quad (33)$$

where λ_{gm} is the guide wave length evaluated at $\lambda = \lambda_m$.

Substituting the measured value of slope and the Y_o computed from the dimensions of the cavity into Eq. (33), permits the evaluation of L_m/C_m. Therefore, L_m and C_m can be determined because their product is known from the anode resonant frequency. The parameters of the iris can be obtained from its dimensions or they can be determined with the aid of additional points on the tuning curves, preferably at frequencies far removed from that of an anode resonance. Tuning curves computed from equivalent circuits give excellent agreement with experimental results.

Stabilization Factor. The energy stored in the tuning cavity of a magnetron partially stabilizes it against frequency changes resulting from variations in rf load and input conditions and is an important consideration that must be taken into account to arrive at a wide-tuning-range design. The division of stored energy in the system is given by the stabilization factor S,[14] which is defined as the ratio of the total energy stored in the magnetron anode, coupling iris, and tuning cavity to that stored in the magnetron anode alone.

$$S = \frac{W_m + W_i + W_c}{W_m} \quad (34)$$

An expression for this quantity in terms of the parameters of the equivalent circuit can be obtained by use of circuit theory. The theory states that in a resonant system the stored energy is equal to the peak instantaneous energy in either the electromagnetic field or the electrostatic fields. At a frequency other than resonance, these energies are not equal; the stored energy in the system is always the smaller of the two. It has also been shown[15] that, in general,

$$B_s = \frac{2\,(W_e - W_h)\,\omega}{|E|^2} \quad (35)$$

$$\frac{dB_s}{d\,\omega} = \frac{2(W_e - W_h)}{|E|^2} \quad (36)$$

where B_s is the shunt susceptance, E is the terminal voltage, and W_e and W_h are the peak instantaneous stored electric and magnetic energies in the system.

Solving Eqs. (35) and (36) for W_e and W_h and taking the smaller value to be the stored energy W_s gives

$$W_s = \frac{|E|^2}{4}\left(\frac{dBs}{d\omega} - \left|\frac{Bs}{\omega}\right|\right) \quad (37)$$

Combining Eqs. (34) and (37) gives

$$S = 1 + \frac{\left[\frac{dB_c}{d\omega} + \frac{dB_i}{d\,\omega}\right] - \left[\left|\frac{B_c}{\omega}\right| - \left|\frac{B_i}{\omega}\right|\right]}{\frac{dB_m}{d\,\omega} - \left|\frac{B_m}{\omega}\right|} \quad (38)$$

This expression is valid for computing stabilization at all values of λ except at $\lambda = \lambda_m$, where B_m and B_c go to infinity.

An expression for S valid at $\lambda = \lambda_m$ can be obtained from the following general network relations:[15]

$$X = \frac{2\,(W_h - W_e)\,\omega}{|I|^2} \quad (39)$$

$$\frac{dX}{d\,\omega} = \frac{2\,(W_h - W_e)}{|I|^2} \quad (40)$$

where X is the series reactance and I is the terminal current.

Solving for W_h or W_e and evaluating at resonance gives the stored energy. Substituting this value of stored energy into Eq. (34) gives the following expression for S which is valid $\lambda = \lambda_m$

$$S/_{\lambda=\lambda_m} = 1 + \frac{\left.\frac{dX_c}{d\,\omega}\right|_{\lambda=\lambda_m}}{\left.\frac{dX_m}{d\,\omega}\right|_{\lambda=\lambda_m}} \quad (41)$$

The stabilization factor can be computed at any point on the tuning curves by use of Eqs. (38) and (41) and the expressions for B_m, B_i, and B_c.

It has been shown that the speed of starting in a particular mode of oscillation varies inversely with the degree of stabilization of the mode.[16] Therefore, if it is desired to have a broadband tunable design which will operate at fast rates of voltage rise and without the added complexities of damping the interfering modes, it is necessary that the stabilization factor be small. Furthermore, from considerations of Eqs. (32), (38), and (41), it can be shown that, once the maximum value of stabilization factor has been fixed, the greatest tuning

range of a particular mode is obtained by a tight coupling between anode and tuning cavity.

Mode Selection Criterion. It would be very desirable to have a criterion that would enable one to predict the exact operating frequency range of a tunable magnetron. This range, however, is difficult to establish because mode selection is a very complicated process involving many variables. Some of these variables include power supply impedance, cathode size and emission, field pattern distortion, voltage pulse rate of rise, mode separation, stabilization factor, and rf load.

An approximate criterion useful for maximizing the tuning range can be established[17] from the principle of minimum dissipation proposed by B. Von der Pol.[18] This principle states that if a self-oscillatory system has more than one possible mode of operation, it will operate in the mode requiring a minimum dissipation from the power sources present in the system. It follows that a magnetron will tend to operate in the mode that has the greatest ratio of energy stored in the magnetron anode to the total energy dissipated, i. e., it will tend to operate in the mode with the highest value of Q_L/S.

That this ratio is not the only factor that determines the mode selection is evident from an inspection of the tuning curves shown in Fig. 22. Here it is seen that one of the (N/2)-1 modes does not tune and therefore has a stabilization factor approximately equal to unity. Furthermore, because this mode has two-fold symmetry, the weak coupling to the tuning cavity indicates that it will also be lightly coupled to the output by a transformer diametrically opposite the tuning cavity. Even though this mode has a high value of Q_L/S, tubes operate in the π mode with relatively little mode interference over at least a part of this region. However, it is known that this non-tunable mode does cause serious interference in some regions and that decreasing its Q_L/S decreases the interference.

Four-Cavity Tunable Magnetrons[19]

The advantages accrued by coupling tuning cavities to resonators (see Fig. 23) can be explained by showing how this arrangement overcomes the limitations encountered with single-cavity tuning. It is evident from the symmetry and number of tuning cavities that the field pattern will be less distorted. The tuning curve that passes through the π-mode wave length at $d = \lambda_g/2$ will also pass through the (N/2)-2 mode wave length at $d = 0$; the result is a relatively large frequency range over which the π-mode configuration is only slightly distorted. Tests on a pulsed, high-power, X-band tube have shown that there is no appreciable degradation in performance from pattern distortion for an 18 per cent frequency band, this percentage being 1.2 times the separation between the π mode and the (N/2)-1 mode. The uniformity of output coupling as a function of frequency is also improved, because the impedance of the output resonator remains more nearly constant. No compensation for changes in coupling is required when the tube is tuned.

Single-cavity tunable tubes, as shown in Fig. 20, have one (N/2)-1 mode that does not couple to the tuning cavity and output wave guide. As indicated previously, this fact establishes a favorable condition for oscillation in this mode. A magnetron with four cavities equally tuned and symmetrically arranged with respect to the anode does not have this difficulty. Both (N/2)-1 modes have high tuning rates because their doublet nature requires that thay have positions for maximum coupling at resonators separated spatially by 90 degrees. In fact, for a tube with 16 resonators as shown in Fig. 23, both doublets of all odd-numbered modes will tune. Furthermore, the loading of the (N/2)-1 modes by the output can be approximately equalized over a wide frequency range by locating the output at 45 degrees with respect to the tuning cavities and by tuning two diametrically opposite cavities so that they are slightly longer than the other two cavities. This arrangement introduces a discontinuity which orients the (N/2)-1 modes for equal coupling to the output resonator. Thus, the Q_L/S ratio can be made much more favorable for the desired mode with four-cavity tuning than by use of a single-cavity tuning. It is also evident that a minimum of four cavities must be tuned to satisfy the symmetry requirement and provide, at the same time, tuning curves with favorable Q_L/S ratios.

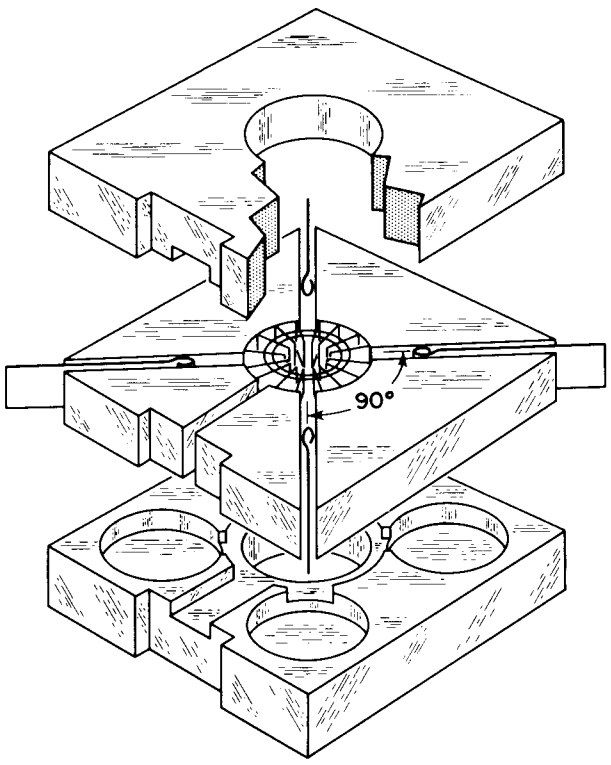


Figure 23. Magnetron with Four Tuning Cavities

Four-cavity X-band pulsed magnetrons of the type shown in Figs. 23 and 24 give tuning ranges of 1100 megacycles or greater (as compared to 300 megacycles for single-cavity tuned tubes).

MAGNETRON CATHODES

The cathode is subject to a number of stringent conditions which must be met for proper magnetron operation.

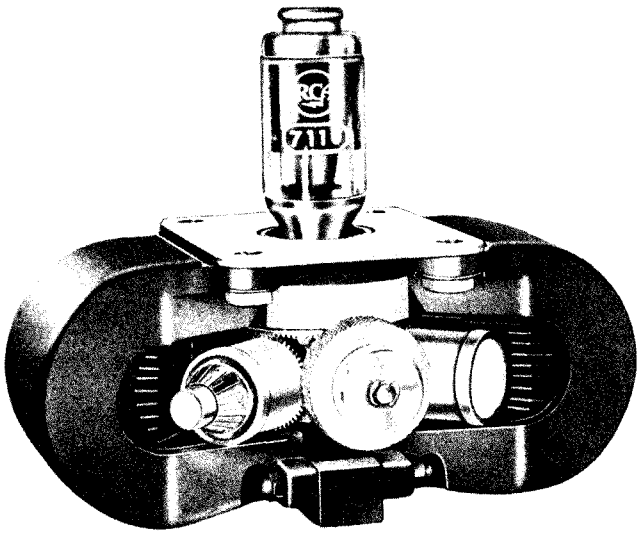


Figure 24. Tunable Magnetron, RCA-7110

First, the cathode must furnish a large current density without sparking or becoming temperature limited to an extent that sufficient change in potential distribution occurs to cause moding or even a decrease in efficiency. Pulsed tubes usually operate at current densities of 25 amperes per square centimeter and values as high as 100 have been attained.

Second, the cathode must withstand bombardment by returned electrons and ions and must have sufficient thermal capacity and cooling to dissipate the back bombardment energy without overheating. Often it is necessary to decrease the heater power to zero when a magnetron is oscillating. The cathode diameter for a given anode design is limited by the decrease in electronic efficiency which occurs if the cathode is made too large.

Third, the cathode must provide an axial boundary to electron flow. The magnetron operates at a voltage several times that of a tube with similar geometry but with no magnetic field. The resulting high space-charge repulsion forces would cause an excessive axial current from the ends of the interaction space if no boundary were provided.

In view of these conditions, it is not surprising that the cathode is usually the element responsible for determining the end of life of the magnetron. As a corollary, it follows that the design of a long-life magnetron requires a massive cathode and operation of it at low current density and at a low temperature. This design has been used in the case of the type 6521 magnetron (used in commercial airline weather radar) to give tubes with a mean life of 6000 hours as compared to the life of 250 to 500 hours for magnetrons less conservatively designed.

The large current density required from a pulsed magnetron cathode is not such a problem as it might initially appear. First, pulsed magnetrons are normally operated at short pulse lengths — a few microseconds, or even less, for high-frequency tubes. The peak pulse emission which can be drawn from a cathode is many times the steady-state emission. In addition, because most cathodes are reasonably good secondary emitters, the back-bombarding electrons knock out secondaries which add to the primary emission.

The cathode thermal dissipation problem is more serious. Three to ten per cent of the input power may be expected to appear as cathode back-bombardment power. Maintaining a reasonable cathode temperature often requires the use of heavy support structures and large radiating surfaces to increase the thermal dissipation. The extension above the top end hat of the center cathode in Fig. 25 is designed to increase the area available for thermal radiation. The roughening of surface areas of the cathode support by sand blasting or spraying with molybdenum powder is a means of increasing thermal dissipation.

The increased thermal dissipation required to compensate for back bombardment introduces the additional problem of heater life. The available space for the heater is very limited and, therefore, magnetron heaters often must be operated at temperatures far higher than those employed in other tube types. The high temperatures complicate the task of insulating the heater; consequently, heater shorts due to the failure of the alumina oxide insulation are a problem. Additions of small amounts of titanium dioxide and chromium to the coating are sometimes made to increase the thermal emissivity of the heater so that its temperature for a given amount of heat radiated is decreased. In some instances, a molybdenum powder coating is sintered to the heater before coating it with insulation to increase the adherence of the insulating coating. Quality control measures are also indicated; the cycling of sample heaters out of each coated lot in dummy cathodes is frequently done before the lot is approved for use.

The normal method of providing axial boundaries to the space charge is by the use of enlarged diameters of the cathode support at each end of the cathode. These supports are called end hats or end shields. Here again, another difficulty is introduced by this solution. If the end hats have large electron emission, either primary or secondary, they are useless. The most effective techniques for reducing end-hat emission are the prevention of contamination by the emission material and the design of the cathode structure to provide low end-hat temperatures by isolation of the end hats, or by other means. In some cases, these methods are not adequate and it is therefore necessary to reduce end-hat emission by chromizing (diffusion of chromium into molybdenum end hats and then oxidation of the material) or by coating with zirconium powder. Both of these techniques reduce the end-hat temperature by increasing radiation; their effectiveness is partly due to this reduction.

Several typical magnetron cathodes are shown in Fig. 25. The left cathode, of the thoria-dispenser type, is for a 200-watt X-band CW tube. The outer sleeve is made of molybdenum and has a number of 0.010-inch holes drilled in it. The inside of the cylinder is packed with thoria, some of which reduces to thorium and migrates through the holes to the outside

of the cathode. The cathode operates at a temperature of 1300 to 1400 C and gives a CW current density of 1.5 amperes per square centimeter. Its life is relatively short (a few hundred hours) due to the high operating temperature and the small size which results in a mechanically weak structure. Larger cathodes of this type give lives of a thousand hours or more at similar current density and temperatures.

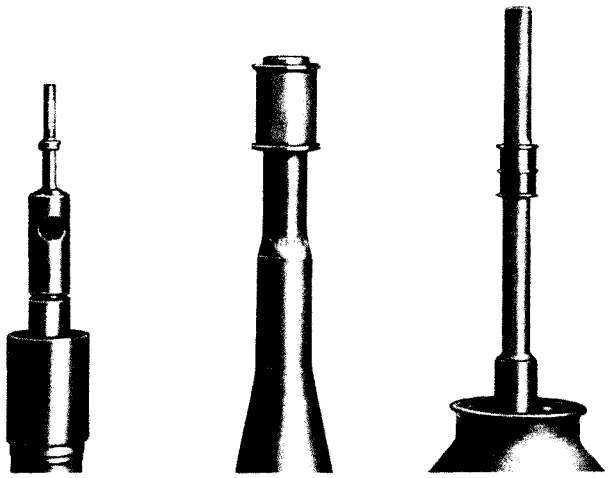


Figure 25. Typical Magnetron Cathodes

The cathode in the center is for the type 6521, an 80-kilowatt pulsed C-band magnetron which was designed specifically for long life. It is of the sintered nickel type. A layer of fine nickel (0.030-inch thick) powder is pressed on the inner molybdenum sleeve and sintered in a hydrogen furnace. A barium carbonate emission mixture is cataphoretically coated on the sintered nickel, filling the pores to a considerable depth. The nickel is then compacted in a press to the final cathode size. This cathode operates at a temperature of about 850 C at a current density of 5 amperes per square centimeter for pulsed conditions and gives a mean life in excess of 6000 hours.

The cathode on the right in Fig. 25 is used in the 7008/7010/7111 series of tunable 200-kilowatt X-band magnetrons, one of which is shown in Fig. 24. This cathode has an active surface of type-A barium aluminate impregnated tungsten. It consists of a sleeve of porous tungsten impregnated with barium aluminate and brazed to the molybdenum supporting structure. The end hats are chromized to prevent emission. The cathode operates at a temperature of 800 to 900 C and a current density of 27 amperes per square centimeter. The life is of the order of 1000 hours. These results demonstrate the value of the impregnated-tungsten cathodes because the untuned version of this tube (4J50) which used a sintered nickel cathode had a life of only about 250 to 500 hours.

OUTPUT CIRCUITS

Most magnetrons employ one of two major types of output circuits. At frequencies below 400 megacycles, low- and moderate-power magnetrons usually have a coaxial-line output circuit. This circuit consists of a loop coupled to the rf magnetic field in the inductive region of one of the resonant cavities and of the magnetron anode. The loop is connected to a coaxial line which is brought outside the vacuum-tight envelope through a glass or ceramic seal. The coupling coefficient can be crudely approximated by the ratio of the area of the loop to the area of the inductive region of the cavity (Ref. 5, p. 191).

Magnetrons for frequencies above 4000 megacycles, as well as magnetrons designed to operate at very high power at lower frequencies, usually have a wave guide output. This type of output normally consists of a coupling slot in the back of one of the resonators, followed by a $\lambda/4$-ridge-wave guide matching transformer which connects to a standard wave guide. A matched glass or ceramic window in the wave guide provides the vacuum enclosure. The impedance of the ridged-wave guide matching section is the geometric mean between the impedance at the back of the resonator and the impedance of the standard wave guide. Because the impedance at the back of the resonator is normally in the range of a fraction of an ohm to a few ohms, the matching section impedance is roughly 40 ohms to give a match to the 377-ohm standard wave guide impedance. The length of the matching section must be near enough to $\lambda/4$ so than an excessive amount of reactance is not coupled into the tube. Other than causing a frequency change, the detrimental effect of coupling reactance into the tube is one of pattern distortion of the rf field patterns around the anode. The usual criterion for coupled reactance is that is does not have a value in excess of the coupled resistance. This requirement is not a stringent; the simple $\lambda/4$ transformer satisfies it over a wide enough bandwidth for even the widest-tuning-range tunable magnetrons.

The degree of coupling of the magnetron to the load is indicated by the external Q which is defined as the energy stored in the resonant circuit per cycle divided by the power to the load. Q_{ext} may be measured by cold impedance measurements or may be calculated approximately.[5] The pulling figure of the magnetron is defined as the maximum frequency change occurring when a load having a voltage standing-wave ratio of 1.5 is varied through all phases. As Q_{ext} is decreased, the circuit efficiency and the pulling figure increase. The value of Q_{ext}, therefore, is a compromise between optimum efficiency and the value of pulling figure which can be tolerated. A high pulling figure may cause long-life effects or it may cause the magnetron to mode.

The Rieke Diagram[5] shows the performance of the magnetron with constant input conditions but with varying load. It consists of contours of constant frequency, power output, and voltage (taken at constant anode current) plotted on the Smith impedance chart.

THE MAGNETIC CIRCUIT

There are two aspects to the design of magnetic circuits for magnetrons: (1) the design of the pole-piece geometry to give the desired shape of magnetic field in the interaction space, and (2) the design of the per-

manent magnet (or in some cases, an electromagnet) to furnish the required flux and magnetizing force to the pole pieces.

The second aspect, the design of the permanent magnet, is straightforward except for the necessity of using empirical factors to allow for the flux leakage along the magnetic circuit. Literature concerning permanent-magnet design, including estimates of leakage factors, can be obtained from most manufacturers of permanent magnets. Due to the great important of the leakage factors, however, it is wise to entrust the design of the permanent magnet to a manufacturer of permanent magnets whose experience enables him to make a more accurate determination of the empirical leakage factors.

The design of the pole pieces affects magnetron operation (through the resulting field shape) as well as the mechanical design of the tube structure and, therefore, their design must be the concern of the magnetron designer. An axial uniformity of magnetic field to better than 5 per cent over the anode length is desirable. An antibarreling field is especially to be avoided, because it will concentrate the current at the ends of the anode where it is least effective. An antibarreling field also complicates the end-hat problem because it tends to allow the electrons to escape past the end hats.

The magnetic field shape of a given set of pole pieces may be determined through the use of the electrostatic plotting tank or by the curvilinear square method of field plotting. Both of these techniques require that no part of the pole piece become saturated. There are various forms of magnetometers available that are small enough to permit experimental measurement of the field shape.

In some cases it may be desirable to combine the design of the magnetic field shape at the ends of the anode with the end-hat design. The objective is to obtain radial rectilinear electron flow with a sharp boundary at the end of the anode. Plotting tank techniques may be employed to determine the end-hat and pole-piece shapes which give magnetic and electrostatic field shapes that combine to give the desired electron flow.

CONSTRUCTION TECHNIQUES

Cavity magnetrons require a considerable number of machined parts. The anode parts are usually made of oxygen-free high-conductivity copper (OFHC), a material not the most desirable for easy machining. It is not surprising, therefore, that most of the special construction techniques developed for magnetrons involve machining methods or other techniques for making precision parts.

The anode is usually the most complicated magnetron part, as well as the one requiring highest precision. Hole and slot anodes of the type shown in Fig. 2A (usually used for frequencies below 4000 megacycles) are usually made by jig-boring the holes and either slotting or broaching the slots. Tolerances of one-half to one mil are normally maintained on the diameter of the holes and on the width of the slots.

The double-ring, strapped-vane anode shown in Fig. 16 is the prevalent type for frequencies up to 10,000 megacycles. It is normally constructed by machining or stamping the vanes and straps and precision brazing them in an anode shell. This structure, requiring close tolerances, is probably one of the most complicated types made by precision brazing. Stainless-steel jigs, with a thermal expansion matching that of copper, are oxidized in a wet hydrogen atmosphere to acquire a surface which will not adhere to the parts and which will not be wet by the brazing material. The parts are assembled in the jig together with the brazing material, usually silver-copper eutectic or a gold-copper alloy, and the braze is made in a hydrogen furnace. Tolerances of about one mil are maintained on the spacings between vane tips and the spacings between the straps.

The rising-sun type of anode, shown in Fig. 18, is used in very high frequency tubes — 10,000 megacycles and higher. It is almost always made by hobbing, although for lower frequency operation anodes are occasionally made by precision brazing. Hobbing is the technique of pressing a hardened steel mandrel (a male master plug called a hob) into a metal blank to produce the desired interior configuration. The complete rising-sun anode is made by this simple process except for machining the ends of the anode. The force required to hob an anode varies from a few tons for small high-frequency anodes to over 50 tons for larger anodes. Hot hobbing is sometimes used to reduce the force required or to permit hobbing of anodes with a large ratio of radial vane length to vane thickness. The copper is maintained in the annealed state by heating it to about 550 C during the hobbing operation. Anodes with vanes of 0.020-inch radial length and 0.004-inch thickness have been hobbed by this technique. Annealing the copper before hobbing is of little or no value because it immediately hardens as it starts to flow. The precision of the anode depends primarily on the precision with which the hob is ground. The hob must be ground after the steel is hardened, because steel distorts enough during hardening to make it impossible to maintain the required tolerances. For very high frequency tubes, tolerances are in fractions of a mil.

High-power pulse magnetrons sometimes require anode dissipation densities so great that the surfaces of the vane tips reach temperatures above the melting point of copper during the pulse. For short pulse lengths, analysis shows that the temperature gradient occurs almost entirely in the first mil of metal thickness at the surface. Use of molybdenum or tungsten vanes does not help because the steady-state temperature gradient along the length of the vane is so much higher than that of copper that it overbalances the benefit of the higher melting temperature of molybdenum or tungsten. The use of copper vanes tipped with a thin layer of tungsten, however, permits peak anode dissipation densities several times that of vanes made entirely of copper. Tubes with tungsten-tipped vanes have been made by brazing a strip of tungsten 0.010 to 0.015-inch thick to the vane face. The difference in expansion is handled by maintaining radial pressure of the vane tip so that radial tolerances are held. As the

assembly cools, the copper deforms in a longitudinal direction which does not affect tube operation adversely.

Due to the high thermal gradient in the first mil of thickness of the tungsten there are stresses of large magnitude repeated millions of times per hour of operation. Extrapolation of theory indicates the surface of the tungsten should disintegrate. However, the limited life experience available to date has not indicated any such problem, possibly because of the short duration of the stresses.

Further discussion of magnetron construction techniques is given in Ref. 5.

MAGNETRON PROCESSING

Magnetrons with a surprisingly poor vacuum can still oscillate. A gassy tube can frequently be aged in, due to the large internal metal area available for gettering. Often, a tube will start oscillating before the glow discharge occurring during aging disappears. During operation, however, even a small amount of gas has detrimental effects. Moding on the leading edge of the pulse is very sensitive to poor vacuum; the tube arcs excessively and the spectrum is poor. Furthermore, a tube with a poor vacuum must be seasoned every time it is operated. High-power pulsed magnetrons have high voltage gradients between electrodes, usually in excess of 500 volts per mil. It is important, therefore, to obtain a good vacuum.

The degree of processing required to achieve a good vacuum depends to a considerable extent upon the type of tube and the requirements placed on it. Tubes having high current densities and high-voltage gradients, as well as those required to meet very exact specifications on such matters as moding on the leading edge of the pulse, require very stringent processing.

Pre-exhaust Treatment

The maintenance of strict cleanliness during assembly is, of course, a minimum requirement of pre-exhaust treatment. Vacuum-firing of cathodes and other metal parts which operate at high temperatures is often necessary. Most magnetrons contain internal pole pieces of cold-rolled steel or soft iron, materials which are notorious for giving off gas. It is usually necessary either to fire these parts in a vacuum or to plate them with nickel and copper. It is important to flow the copper to close any pores in the metal surface.

Exhaust

During exhaust, it is desirable that all metal parts be heated for some time to temperatures in excess of those which occur in operation. However, it may not always be possible to meet this requirement due to the temperature limitations of other parts such as glass seals. Because these limitations restrict the temperature to about 425 C, it has been necessary with some types to increase bakeout schedules to 10 or 12 hours. Another technique is to place a gas burner under the anode during exhaust to increase the anode temperature without causing damage to glass parts. The replacement of glass by ceramic permits an increase of the bakeout temperature to about 500 to 550 C (with a corresponding reduction of bakeout time) in tubes using silver brazing alloys. If all silver is eliminated from the tube, even higher exhaust temperatures are practical. In some cases, particularly in high-power CW or very high-power pulse tubes, it is necessary to operate the tube as an oscillator during exhaust to be sure that all parts are heated to a sufficiently high temperature. Precautions should be taken with such tubes to prevent any material feeding back from the vacuum pump. For this reason, copper foil traps and liquid-air traps are indicated.

Getters

Most magnetron manufacturers do not employ getters; RCA regularly uses them. It is believed that the long life and the ability of RCA magnetrons to meet very exact moding specifications result to a considerable degree from the use of getters. Indications are that getters improve both shelf life and operating life. They also reduce the amount of aging time required in some types.

At present, a flashed barium getter is employed. It is placed in the cathode stem with a shield to prevent flash from reaching the high-voltage region of the tube. Other getters of the bulk type, such as Ceralloy or sintered zirconium powder, are being evaluated.

Aging

Maximum stability, both with respect to arcing and moding, requires that magnetrons be aged at peak currents that are considerably in excess of their operating currents. Aging a tube at currents of one and one-half to two times the operating current is done at a reduced duty factor.

LIST OF SYMBOLS

Note: MKS units are employed throughout this article except where specified otherwise in the text.)

a_1	a function of r_a/r_c (Fig. 9)
b	reduced dc magnetic field
B	dc magnetic field
B_c	susceptance of tuning cavity
B_i	susceptance of coupling iris
B_m	susceptance of magnetron
B_s	shunt susceptance
$\mathfrak{B}$	characteristic dc magnetic field
c	velocity of light
C	capacitance
C_s	strap capacitance
d	length of tuning cavity (Fig. 20C)
e	electronic change
E	rf voltage
f	frequency
F	force (on electron)
g	load conductance presented to magnetron
g_o	load conductance resulting in maximum power output

G_1	magnetron slot conductance
$\mathcal{G}$	characteristic conductance
h	height of magnetron anode
i	reduced anode current
I	anode current (referring to magnetron)
I	rf current (referring to tuning cavity)
$\mathcal{I}$	characteristic current
J_c	cathode current density
k	Hartree mode number
L	inductance
m	electronic mass
n	mode number of resonator system (referring to magnetron)
n	zero or positive integer (referring to tuning cavity)
N	number of magnetron resonators
p	reduced generated power
P	generated power
P_o	output power
$\mathcal{P}$	characteristic power
q	any integer
Q_{ext}	external Q (magnetron to load)
Q_{ext^m}	pseudo external Q (magnetron to tuning cavity)
Q_L	loaded Q
Q_o	unloaded Q
r_a	anode radius
r_c	cathode radius
s	radial distance to back of resonator (Fig. 19)
S	stabilization
T	period of rf oscillation
v	reduced anode voltage
v_e	velocity of electron
V	anode voltage
V_c	Hull cutoff voltage
$\mathcal{V}$	characteristic voltage
w	vane thickness
W	instantaneous stored energy
X	reactance
Y_o	characteristic admittance
β	Langmuir's function for cylindrical diode
ϵ_o	permittivity of free space
η	over-all efficiency
η_e	electronic efficiency
η_c	circuit efficiency
σ	ratio of cathode radius to anode radius
λ	wave length
μ	relative slot width
ω	angular frequency

REFERENCES

1. Hull, A. W., "The Effect of a Uniform Magnetic Field on the Motion of Electrons Between Coaxial Cylinders," Phys. Rev., Vol. 18. p. 31, July 1921
2. Langmuir, I. L., and K. Blodgett, "Currents Limited by Space Charge Between Coaxial Cylinders." Phys. Rev., Vol. 22, p. 347, October 1923
3. Allis, W. P., "Electron Orbits in the Cylindrical Magnetron," MIT Radiation Laboratory Report 9S
4. Hartree, D. R., CVD Magnetron Report No. II
5. Collins, G. B. (ed.), Microwave Magnetrons, MIT Radiation Laboratory Series, Vol. 6, McGraw-Hill, New York, 1948
6. Code M1
7. Clogston, A. M., "Magnetron Scaling Formulae, MIT Radiation Laboratory Report No. 52, January 21, 1943
8. Smith, W. V., "Practical Considerations of Magnetron Design," MIT Radiation Laboratory Report 52-7, August 22, 1943
9. Collins, G. B. (ed.), Microwave Magnetrons, pp. 561-591
10. Herlin, M. A., "Cold Resonance Theory of the Waveguide Tunable Magnetron," Radiation Laboratories Report No. 445, October 1943
11. Smith, W. V., "Magnetron Tuning and Stabilization," Radiation Laboratories Report No. 567, July 1944
12. Jenny, H. K., "A 7000-Megacycle Developmental Magnetron for Frequency Modulation," RCA Review, Vol. XIII, No. 2, pp. 202-223, June 1952
13. Slater, J. C., Microwave Electronics, pp. 84-88, D. Van Nostrand, 1950
14. Collins, G. B. (ed.), Microwave Magnetrons, p. 724
15. Montgomery, C. G., R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits," Radiation Laboratory Series, Vol. 8, p. 135, McGraw-Hill, New York, 1948
16. Collins, G. B. (ed.), Microwave Magnetrons, p. 632
17. Code M2
18. Von der Pol, B., "Oscillation Hysteresis in a Triode Generator with Two Degrees of Freedom," Philosophical Magazine, Vol. XLIII, p. 717, April 1922
19. Code M3

Traveling-Wave Tube Design

E. E. Bliss

Harrison

The term traveling-wave tube covers a broad field of apparently unrelated tube types. Fig. 1 shows the various divisions and subdivisions of the field as well as several specific examples of traveling-wave-tube types. The two main divisions are forward-wave traveling-wave tubes and backward-wave traveling-wave tubes. In the forward-wave type, the electron beam and the wave on the slow-wave structure both travel in the same direction and usually at approximately the same velocity. In the backward-wave type, the beam travels in one direction and the power flow on the slow-wave structure travels in the opposite direction. Both types can be used as amplifiers or oscillators. The two main types can be further subdivided into "O"-type tubes in which the rf energy is extracted from the kinetic energy of the electrons, and the "M"-type or cross-field device in which the energy is extracted from the potential energy of the electrons.

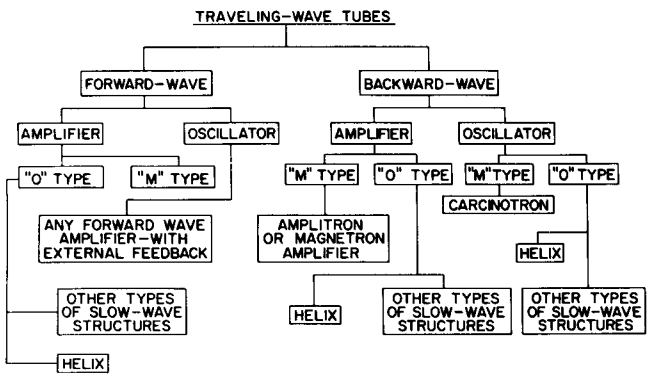


Figure 1. The Various Types of Traveling-Wave Tubes are Divided into Two Basic Classes: Forward-Wave and Backward-Wave Tubes

Comprehensive coverage of all traveling-wave-tube types is beyond the scope of this chapter. However, the fundamental theory for all types is based on the forward-wave traveling-wave-tube amplifier and much of the basic design information for that type is applicable to the other types. Because it is also by far the most common traveling-wave device and is the type most likely to be encountered in actual design practice, the major portion of this chapter is devoted to forward-wave amplifiers.

Where there is thorough treatment in the literature of the design aspects of a specific component, the article is referenced, and this text includes just a brief description of the material covered. However, where the traveling-wave-tube component is usually scaled from a known component, a fuller discussion, scaling factors, and examples of performance are given.

Traveling-wave tubes lend themselves to a fairly accurate theoretical analysis, especially in cases where small-signal approximations are valid. Considerable theoretical work on these devices has been published, and fairly detailed design formulas are available. The work of J. R. Pierce,[1] in particular, has long served as a reference; the nomenclature and symbolism used in this book are universally accepted.

The Appendix includes a sample design procedure for a one kilowatt S-Band traveling-wave tube and covers in detail the various design steps that must be followed.

TRAVELING-WAVE TUBE OPERATION

The traveling-wave tube overcomes many of the limitations found in other types of tubes operated over 1000 megacycles. For example, in a triode operated at very high frequencies, the transit time, i. e., the length of time required for an electron leaving the cathode to pass through the grid region and arrive at the plate, can be an appreciable portion of one rf cycle, resulting in low efficiency. At these frequencies the reactance presented by the capacitance of the tube and the inductance of the leads can render the tube inoperable.

FORWARD-WAVE TRAVELING-WAVE TUBE

The limitations of the classical triode are overcome in the forward-wave traveling-wave tube as shown in Fig. 2. The electron beam is preaccelerated in the gun region and then directed, at maximum velocity, through an interaction circuit. A focusing device is used to confine the beam in the interaction region. The rf energy to be amplified enters through the input coupler and travels along the interaction structure at approximately the same velocity as the electron beam. The interaction between the electron beam and the signal on the slow-wave structure occurs continuously down the entire length of the tube. The amplified signal is then extracted by the output coupler and the spent beam continues on to the collector. By maintaining the electron beam and the wave on the circuit in synchronism down the length of the tube, it is possible to take advan-

tage of the transit time of the electrons to extend the interaction between the wave and the beam over a considerable distance.

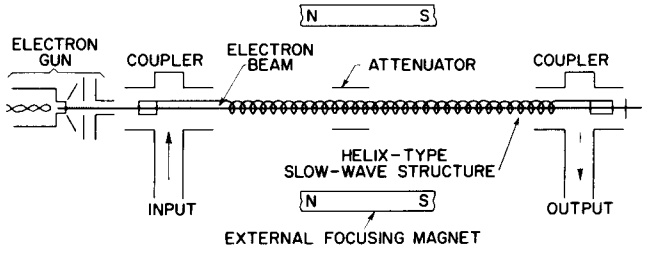


Figure 2. Basic Elements of a Traveling-Wave Tube

The use of a slow-wave structure and the absence of resonant circuits makes possible extremely wide frequency ranges, typically an octave. The gain-bandwidth product of traveling-wave tubes is very high compared to those of lower frequency amplifiers and other types of higher frequency amplifiers using resonant structures. For example, a typical tube covering an octave at a ten-centimeter wavelength and having a power gain of 30 db has a gain bandwidth product of 2×10^{12}.

Fig. 3 shows the interaction between the slow wave structures and the electron beam. The wave travels down the slow-wave circuit; at a given instant, the rf voltage is alternately positive and negative along the length of the slow-wave circuit. Thus, the electrons which are moving inside the slow-wave circuit, at approximately the same velocity as the wave, are affected by the small ac voltages. An electron at point A is subjected to a small positive voltage with respect to the average voltage of the beam and is slightly accelerated. An electron at point B, where the rf field is zero, maintains a constant velocity, while an electron at point C encounters small negative voltage and is decelerated. Thus, as the beam moves through the slow-wave structure, the accelerated electrons catch up with those electrons which have been slightly decelerated and form bunches or groups as they progress through the slow-wave structure. These bunches are actually rf current pulses which can induce a corresponding voltage back onto the slow-wave structure and amplify the original signal. The amplified signal then reacts further on the electron beam and causes even closer and more effective bunch formation. In this way, the signal is amplified on the slow-wave structure until it is finally removed through the output coupler.

Because the rf energy from this interaction comes from the kinetic energy of the electron beam, the beam slows down near the output region of the slow-wave structure. Thus, for optimum interaction and efficiency, the electron beam must travel at a slightly greater velocity than the wave on the slow-wave structure.

For tubes having a gain of more than 10 db, some form of attenuator is required to avoid the amplification of reflections and subsequent oscillation. This attenuator, which is usually some form of loss applied to the slow-wave structure in the center of the tube, effectively cuts off the circuit wave at that point. The signal is carried through the attenuator on the electron beam which has already been partially bunched by the interaction in the input region. The circuit wave is then set up again on the circuit at the output side of the attenuator and the amplification continues in the output section of the tube. However, waves traveling from the output end of the slow-wave structure do not interact with the electron beam and thus cannot be carried past the attenuator.

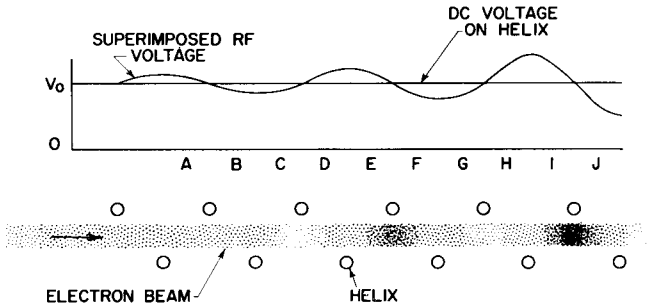


Figure 3. Mutual Interaction between an RF Signal on the Helix and the Electron Beam Results in an Increasing Signal Wave on the Helix

BACKWARD-WAVE TRAVELING-WAVE TUBE

The operation of a backward-wave tube is basically the same as that of the forward-wave amplifier, except that the wave on the slow-wave circuit travels in the opposite direction to the electron beam. The interaction is identical with that in the forward-wave amplifier, but the electrons are made to react as if the wave were going in the same direction as the electrons and at the same velocity as the electron beam in the following manner. It is assumed that the slow-wave structure is a folded waveguide, as shown in Fig. 4. A hole is drilled through the length of the structure as shown so that periodically the electrons are inside and outside the waveguide. The electrons follow a more direct path than the wave in the waveguide and therefore can travel at considerably less than the speed of light and still maintain synchronism. An rf wave travels from the right side of the structure to the left side and at the same time an electron beam travels from left to right. As the electron travels through the waveguide, it encounters only one phase of the rf wave at point A. If the velocity of the electron is arranged so that the same phase of the rf wave is present at the time it enters the waveguide again at point B, the electron will react exactly as it did at point A.

This arrangement has numerous possible feedback paths. For instance, the electron beam is slightly bunched as it passes through point A and, therefore, adds to the rf signal when it arrives at point B. Meanwhile, the rf signal which leaves point A and travels to the left is amplified and causes greater modulation at point C. As the electron beam travels from left to right it becomes more and more heavily modulated and the wave traveling from right to left becomes greater

in amplitude. Thus, at the point at which the wave has its greatest amplitude, the beam is just entering the interaction region and has no modulation. At the point at the right side of the structure at which the greatest bunching has taken place, the rf amplitude is very low. Consequently, the efficiency of the backward-wave oscillator or amplifier is usually quite low. However, because of the feedback nature of the interaction, high gain can be obtained over a relatively short distance. This type of device has a narrow bandwidth because at a given electron velocity, the rf wave is in phase over only a very narrow range of frequencies. On the other hand, the device is readily tunable by varying the electron velocity. Backward-wave oscillators can be made which are tunable over as much as an octave frequency band simply by varying the beam voltage. For this reason, they are quite useful in signal generators and as local oscillators when rapid electron tuning is desired. As amplifiers, they are useful as tunable preselectors.

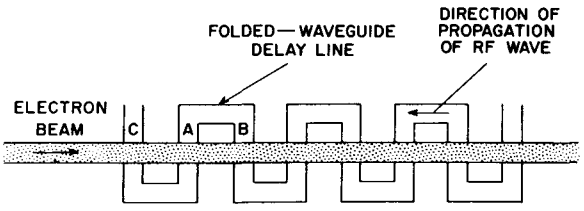


Figure 4. An Electron Beam Propelled through a Slow-Wave Structure in a Direction Opposite to the Propagation of an RF Signal along the Structure Results in Backward-Wave Interaction

SLOW-WAVE CIRCUITS

Many different types of circuits have been used for the interaction structure of traveling-wave tubes. The helix is most commonly used because of its broad bandwidth and ease of fabrication. The primary object of the slow-wave structure is to slow the electromagnetic wave from the speed of light to a velocity at which it can interact readily with the electron beam (usually about one-tenth the speed of light). In the helix, this slowing action is accomplished readily because the wave follows the wire of the helix around a circuitous path while the electron beam travels in a straight line through the center of the helix. In addition, the helix is fairly nondispersive; i. e., the velocity of propagation along the axis of the helix is nearly constant over a wide range of frequencies. Thus, synchronism of the rf wave with the electron beam can be maintained over a wide frequency band. However, the helix has poor heat dissipation properties, and considerable difficulties are experienced with interfering oscillations when operated at very high voltages.

For high-power tubes, the slow-wave structure is usually some form of folded waveguide for backward-wave oscillators or periodically loaded waveguide. If the latter type of structure is used, the loading must be sufficient to slow the wave velocity to the point where it can interact with the electron beam at voltages from 10 to 100 kilovolts. A more detailed discussion of the helix and the other types of slow-wave structures is presented later in this chapter.

ANALYSIS OF OPERATION

The traveling-wave tube is particularly suited to theoretical analysis; a large body of the literature is made up of detailed studies of the small-signal behavior of traveling wave tubes. Because many of the design articles and curves presented in the literature require certain basic background knowledge, this chapter includes the basic approach to small-signal theory, the form of the final equations, graphical use of the final equations, and introduces much of the terminology.

MATHEMATICAL DESCRIPTION OF A WAVE ON A SLOW-WAVE STRUCTURE

Quantitative analysis of the operation of a traveling-wave tube requires a mathematical expression to describe the amplitude and velocity of the rf signal as it travels down the slow-wave structure. The amplitude of an rf signal at a given point as a function of time is expressed as follows:

$$V_{(t)} = V_1 e^{i\omega t} \tag{1}$$

When the wave is constrained to travel in the z direction of the slow-wave structure in a traveling-wave tube, the rf signal amplitude can be defined as a function of both time and distance as follows:

$$V_{(z,t)} = V_1 e^{i\omega t} e^{-i\frac{\omega}{v_p} z} e^{-\alpha z} \tag{2}$$

where v_p, the phase velocity of the wave and α, the attenuation constant are peculiar to a specific slow-wave structure. The first term of this equation shows the phase of the wave as a function of time and thus describes the frequency. The second term shows the phase as a function of distance and thus describes the velocity of the wave as it travels down the slow-wave structure. For positive values of ω/v_p the wave is traveling to the right or for negative values of ω/v_p the wave is traveling to the left. The third term of the equation is a real quantity and thus describes the peak amplitude of the wave as a function of distance. Negative values of α indicate a wave increasing in amplitude while positive values of α indicate a wave decreasing in amplitude. Generally, this equation is written as follows:

$$V_{(z,t)} = V_1 e^{i\omega t - \Gamma_1 z} \tag{3}$$

where

$$\Gamma_1 = i\frac{\omega}{v_p} + \alpha \tag{4}$$

Γ_1 is the propagation constant of a slow-wave structure and fully describes the behavior of an rf signal on any given structure. Once the propagation constant of a structure is determined and an rf signal is fed into the structure, the location and amplitude of the signal can be determined at any time or at any point along the

structure. The propagation constant can be calculated or experimentally determined for any type of slow-wave structure.

MATHEMATICAL DESCRIPTION OF A WAVE ON A SLOW-WAVE STRUCTURE IN THE PRESENCE OF AN ELECTRON BEAM

When an electron beam interacts with the wave on a slow-wave structure, the propagation constant of the structure Γ_1, is modified by the presence of the beam and may be denoted as Γ. Thus, the equation for the wave on the slow-wave structure in the presence of the electron beam becomes

$$V_{(z,t)} = V_1 e^{i\omega t - \Gamma z} \tag{5}$$

This new propagation constant Γ fully describes the amplitude and velocity of the rf signal at any given time and place in the slow-wave structure of an operating traveling-wave tube. Once Γ is determined, the relative signal amplitudes at the input and output of a traveling wave tube can be compared, and the gain calculated.

THE DETERMINATION OF Γ

The development of the equation describing Γ is basically the same as that for the interaction between the wave and the electron of transmission line down which the wave is propagated and that a nearby electron beam has a velocity close to the phase velocity of the wave on the transmission line. A small ac or rf signal is assumed to be on the transmission line and the effect of this signal on the electron beam is calculated. Then it is assumed that a small ac signal is on the electron beam (effectively a small amount of bunching or rf modulation) and the effect that this modulation on the electron beam has on the transmission line is calculated. The two relationships are equated and the resulting transcendental equation defines Γ in terms of the propagation constant of the slow-wave structure and the parameters of the beam.

$$\frac{iKI_0\beta_e\,\Gamma^2\,\Gamma_1}{2V_0(\Gamma_1^2-\Gamma)(i\beta_e-\Gamma)} = 1 \tag{6}$$

where

$\Gamma_1 = -i\frac{\omega}{v_p}$ (it is assumed that the slow-wave structure is lossless)

$\beta_e = \frac{\omega}{u_0}$ (where u_0 is the velocity of the electron beam)

K is the impedance of the slow-wave structure encountered by the beam and may be calculated or determined experimentally;

I_0 is the beam current;

V_0 is the beam voltage.

It is further assumed that v_p is not too different from u_0 and that Γ is not too different from Γ_1. The number of symbols may be reduced by defining C^3 as follows:

$$C^3 = \frac{KI_0}{4V_0} \tag{7}$$

The determinantal equation for Γ may then be expressed as follows:

$$\frac{2iC^3\,\Gamma^2\,\Gamma_1}{(\Gamma_1^2-\Gamma^2)(i\beta_e-\Gamma^2)} = 1 \tag{8}$$

This equation is in the fourth power of Γ and has four possible solutions. The solution is simplified by further assuming that the electron velocity is equal to the cold-phase velocity of the structure; i.e., $-\Gamma_1 = i\beta e$.

The solution has the following form:

$$\Gamma = i\beta_e - \delta\beta_e C \tag{9}$$

where

$$\delta_1 = \frac{\sqrt{3}}{2} - \frac{i}{2}$$

$$\delta_2 = \frac{-\sqrt{3}}{2} - \frac{i}{2}$$

$$\delta_3 = i$$

The fourth solution describes the wave traveling in a direction opposite to the beam, but not interacting with the beam. This solution is lost in making the approximation required to arrive at the other three solutions.

The propagation constant Γ in terms of β_e and Pierce's gain parameter C can be substituted into Eq. (5) to calculate the gain of a traveling-wave tube.

CALCULATION OF GAIN

When the three solutions for Γ are substituted into the wave equation [Eq. (5)], the following equations result:

$$V_{(z,t)} = V_1 e^{i\left[\omega t - \left(\frac{\omega}{u_0} + \frac{\omega}{2u_0}C\right)z\right]}\, e^{\frac{\sqrt{3}}{2}\beta_e Cz} \tag{10}$$

$$V_{(z,t)} = V_1 e^{i\left[\omega t - \left(\frac{\omega}{u_0} + \frac{\omega}{2u_0}C\right)z\right]}\, e^{-\frac{\sqrt{3}}{2}\beta_e Cz} \tag{11}$$

$$V_{(z,t)} = V_1 e^{i\left[\omega t - \left(\frac{\omega}{u_0} - \frac{\omega}{u_0}C\right)z\right]} \tag{12}$$

The first of these three equations describes a wave traveling at a velocity slower than the beam velocity and increasing exponentially in amplitude. The second equation describes a wave again traveling slower than the beam velocity, but with an amplitude which is decreasing exponentially with distance. The third equation shows a wave traveling faster than the beam velocity, but not changing in amplitude.

An input signal which is fed onto the slow-wave structure is divided equally among these three waves. After

traveling only a short distance down the slow-wave structure, the increasing amplitude of the first wave very quickly swamps out the effect of the other two waves which then have an increasingly negligible amplitude compared to the first wave. Thus, the voltage at the output is

$$V_{out} = \frac{V_{in}}{3} e^{\frac{\sqrt{3}}{2} \beta_e C l} \quad (13)$$

where $\sqrt{3}/2\ \beta_e C$ is the attenuation constant from Eq. (10), l is the length of the slow-wave structure and the 1/3 term comes from the input signal being divided equally between the three waves. The power gain in db can now be written

$$G = 20 \log \frac{V_{out}}{V_{in}} = 20 \log \frac{1}{3} + 20 \log_e e^{\sqrt{3}/2 \beta_e C l} \quad (14)$$

or

$$G = -9.54 + 47.3\ C\ N \quad (15)$$

where N, the number of wavelengths in the slow-wave structure, is equal to

$$\frac{l \beta_e}{2 \pi}$$

Eq. (15) was derived with certain simplifying assumptions: a lossless slow-wave structure, a beam velocity equal to the phase velocity of the slow-wave structure, and no space charge parameter. Practical tubes do not satisfy these assumptions, and the gain equation must be modified to take into account the following effects:

1. Lumped attenuation, which is usually introduced to prevent the tube from oscillating.
2. An arbitrary electron velocity.
3. The passive or higher order mode parameter Q.

The more complete gain equation is now written as follows:

$$G = A + BCN - L_{att} \quad (16)$$

where L_{att} is the loss due to the lumped attenuation. The initial loss parameter A and the growing-wave parameter B become functions of the distributed loss, the electron velocity, and the space-charge parameter QC. The passive-mode parameter Q may be found from Fig. 5, where it is plotted as a function of γ_a for different values of beam-to-circuit diameter. The function γ_a, the radial propagation constant, which is the argument for the Bessel functions which describe the fields in the interaction region, are described under SLOW-WAVE STRUCTURES. The function may be defined as follows:

$$\gamma_a = \frac{\omega}{u_o} = \frac{2 \pi a}{\eta_g} \quad (17)$$

Although not exact, this function is a close approximation for tubes operating below 10,000 volts.

Figs. 6 and 7 show the dependence of A, the initial-loss parameter, and B, the growing-wave parameter on the parameter QC. Fig. 8 shows the attenuation or loss of gain which is to be expected with the introduction of the lumped attenuator. This curve is based on the assumption that the slow-wave circuit is severed and a drift tube of length measured by CN is substituted for the circuit, as is very nearly the case when a lumped attenuator with a loss in the order of 20 db per wavelength or more is used. The wave beyond the drift tube has a level with respect to the incident wave shown by the ordinate.

Almost all the information required to calculate the gain of the traveling-wave tube is now available. The initial loss A, the growing-wave parameter B, and the loss due to the attenuator can be obtained from Figs. 5 through 8. The length of the tube N in helix or slow-wave wavelengths can be readily calculated from the following equation:

$$N = \frac{l \beta_e}{2 \pi} = \frac{1\,506}{\lambda_o \sqrt{V}} \quad (18)$$

The only factor remaining to be determined is the impedance parameter K, which is included in the expression for the gain parameter C. How this parameter is determined is described under Helix and Other Slow-Wave Structures because K is necessarily a function of the slow-wave circuit used.

FORWARD-WAVE AMPLIFIER DESIGN

This section deals with the actual design of the forward-wave traveling-wave tube amplifier, specifically, the helix and other types of slow-wave structures, the electron gun, the focusing structure, the couplers, the attenuator, and the collector.

THE HELIX AS A SLOW-WAVE STRUCTURE

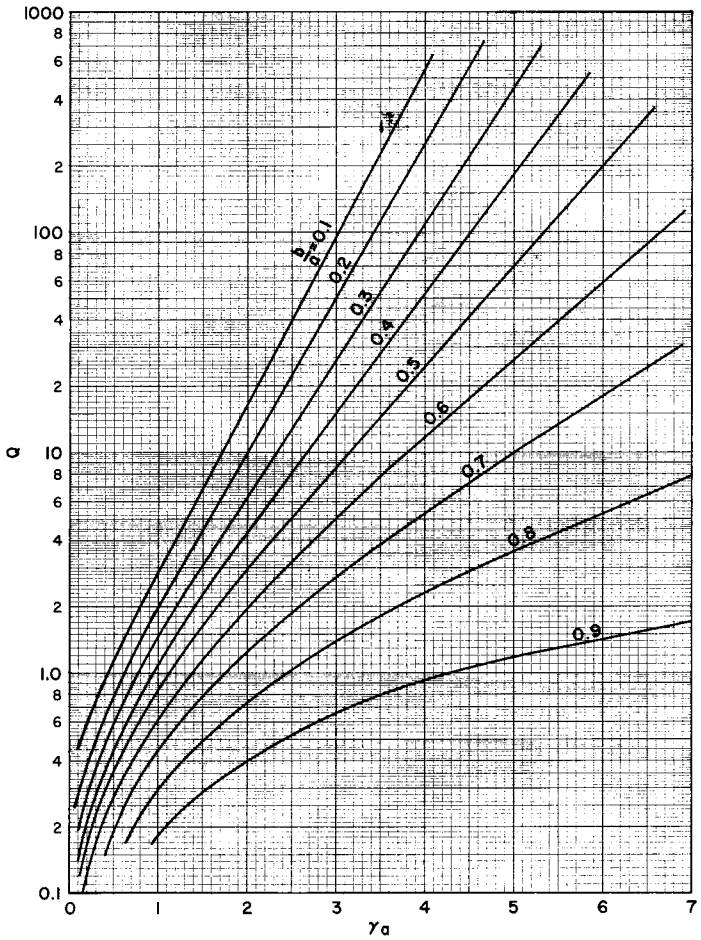
Previously, the expression for the gain of the traveling-wave tube was derived in terms of the initial loss factor A, the slow-wave parameter B, the length of the tube N, and the impedance parameter C. The expressions for the first three of these terms were derived exclusive of the type of circuit used, but the determination of C can be made only from an examination of the slow-wave structure.

Any circuit capable of propagating an rf wave at velocities about one-tenth the speed of light and allowing interaction with an electron beam may be used for the slow-wave structure of a traveling-wave tube. The helix is very nearly ideal for this purpose because it is easy to construct, has a nearly constant phase velocity over a wide frequency range, and can be readily coupled.

The following parameters are used frequently during this discussion:

$$\gamma \simeq \beta = \frac{\omega}{v} \simeq \frac{\omega}{v_p} \quad (19)$$

where γ is the radial-phase constant, β is the axial-phase constant, v is the phase velocity of the wave in



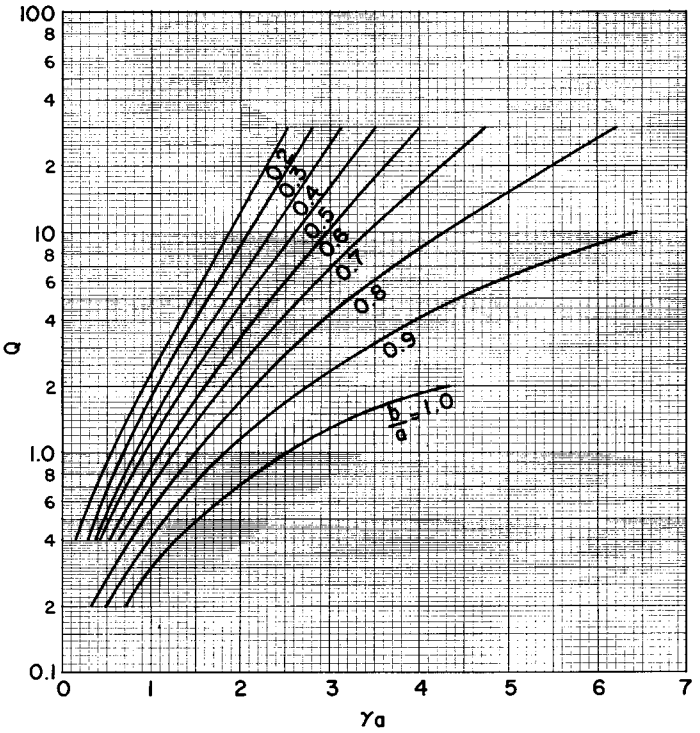


Figure 5. Family of Curves Showing the Passive-Mode Parameter Q as a Function of γa for Different Ratios of the Radius of the Electron Beam to the Radius of the Slow-Wave Circuit (b/a): (A) For a solid beam; (B) For a hollow beam

the presence of the beam and v_p is the phase velocity of the wave on the circuit in the absence of the beam.

$$\beta_o = \frac{\omega}{C} = k \tag{20}$$

which is the free-space phase constant.

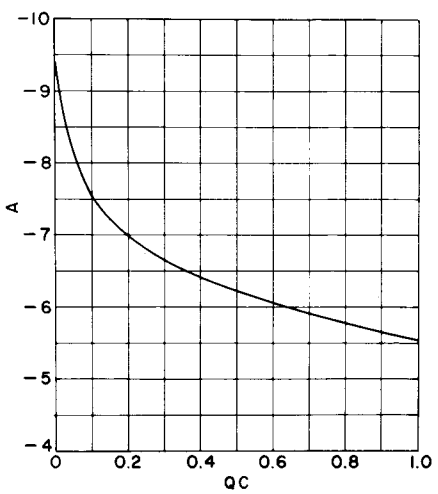


Figure 6. The Initial-Loss Parameter A as a Function of the Space-Charge Parameter QC

The fields inside and outside the helix vary as certain Bessel functions, the argument of which is normally γ_a or γ_r, where r is the radial distance from the axis of the helix and a is the mean radius of the helix. Fig. 9 shows various parameters which describe the helix.

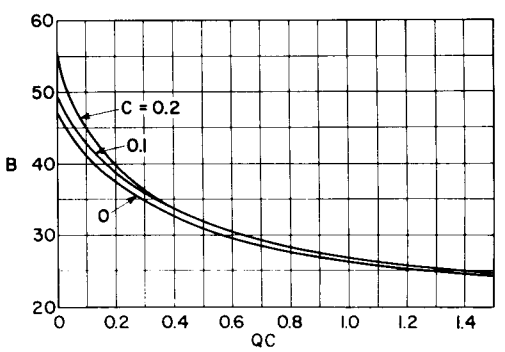


Figure 7. The Growing-Wave Parameter B as a Function of the Space-Charge Parameter QC

Helix Phase Velocity

Fig. 10 shows the axial phase velocity of the helix v_p/c as a function of ka for various pitch angles.

For values of γ_a greater than 1.5, the phase velocity approaches sine ψ and

$$\frac{v_p}{c} = \frac{P}{2\pi a} \tag{21}$$

At lower values of γ_a, the phase velocity varies rapidly with frequency (under this condition the helix is

said to be dispersive), and the proper pitch angle for a reciprocal velocity may be determined from Fig. 10.

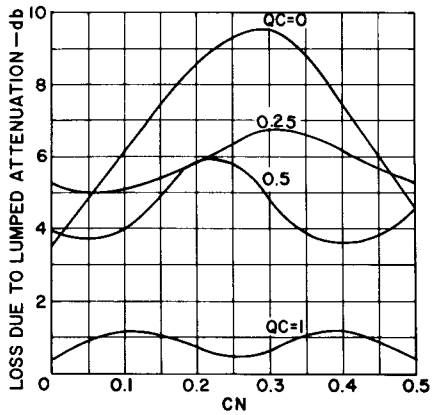


Figure 8. Loss of Gain Resulting from the Use of a Lumped Attenuator for Different Values of the Space-Charge Parameter QC as a Function of the Drift-Tube Length CN

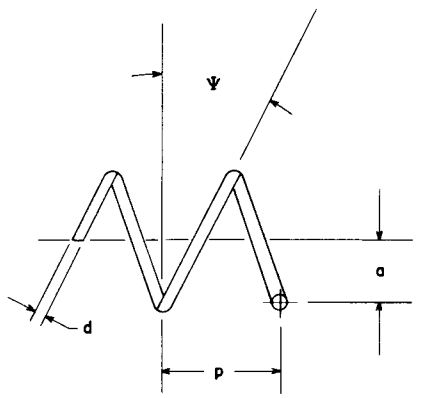


Figure 9. Parameters which Define the Geometry of a Helix

The means by which the helix is supported also affects the phase velocity. The dielectric loading factor DLF is defined as

$$\text{DLF} = \frac{\text{Phase velocity of the helix in the dielectric}}{\text{Phase velocity of the helix in free space}} \quad (22)$$

The DLF for various types of helix supports can vary from 0.7 to 0.9 depending upon the geometrical configuration and material used. Fig. 11 shows the approximate values of DLF obtained for some typical configurations.

Some reduction in dispersion is obtained from the dielectric supports because the DLF is normally not constant with frequency. At higher frequencies, because the fields are closer to the helix and interact less with the dielectric, the DLF decreases, as shown by the typical dispersion curve in Fig. 12. In this case, the small-fluted tubing used to support the helix produced a dielectric loading factor of 0.9 at the high-frequency end and 0.85 at the low-frequency end of the slow-wave structure. As shown in Figure, the configuration resulted in a considerable decrease in dispersion. An equivalent decrease in dispersion with the other types of supports would require a large mass of dielectric near the helix and result in a much greater value of DLF, and would have a detrimental effect on the helix impedance and gain.

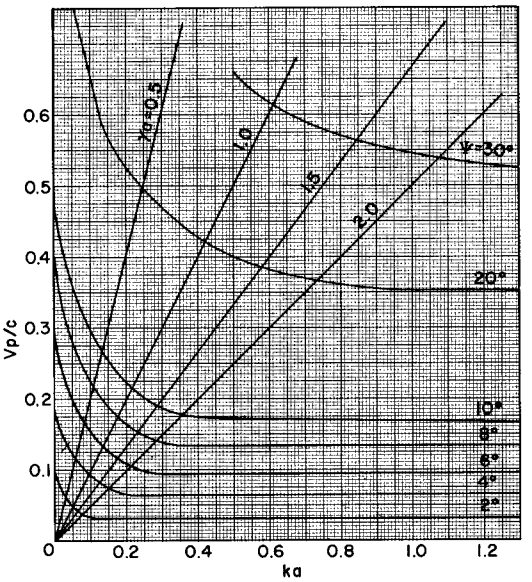
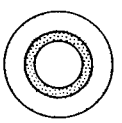
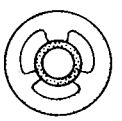


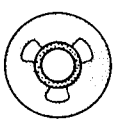
Figure 10. Family of Curves Showing the Phase Velocity of the Helix v_p/c as a Function of ka for Several Values of the Helix Pitch Angle ψ. Lines of Constant γa are Included for Reference.



TYPE OF SUPPORT	DIELECTRIC LOADING FACTOR
SOLID GLASS TUBING	0.70—0.80



FLUTED GLASS TUBING; SMALL FLUTES SMALL CERAMIC RODS	0.85—0.90



FLUTED GLASS TUBING; LARGE FLUTES LARGE CERAMIC RODS	0.75—0.85

Figure 11. Range of Dielectric Loading Factor for Three Types of Helix Support Structures

A conducting shield around the helix may also be used to decrease the dispersion as shown in Fig. 13. Reference 2 fully describes the effect of various size shields on dispersion and on impedance.

Helix Impedance

The axial component of the electric field on the axis

of the helix has an amplitude which varies with γ_a, as shown in Fig. 14A. The impedance K, which is required to determine the gain parameter C [See Eq. (7)], is a function of the axial electric field $E_{(z)}$ and of the phase constant $1/\gamma$. High impedance is desirable for two reasons: first, it enables a higher gain per unit length; and second, it is essential for high efficiency. In general, it has been empirically found that the efficiency η is approximately equal to 2 x C. As shown in Fig. 14A, the electric field on the axis falls off rapidly at higher frequencies, and thus, the impedance is a rapidly decreasing function of frequency. The impedance parameter $F_{(\gamma a)}$ is given as a function of γa in Fig. 15 where

$$K = \frac{[F(\gamma a)]^3 C}{2 v} \qquad (23)$$

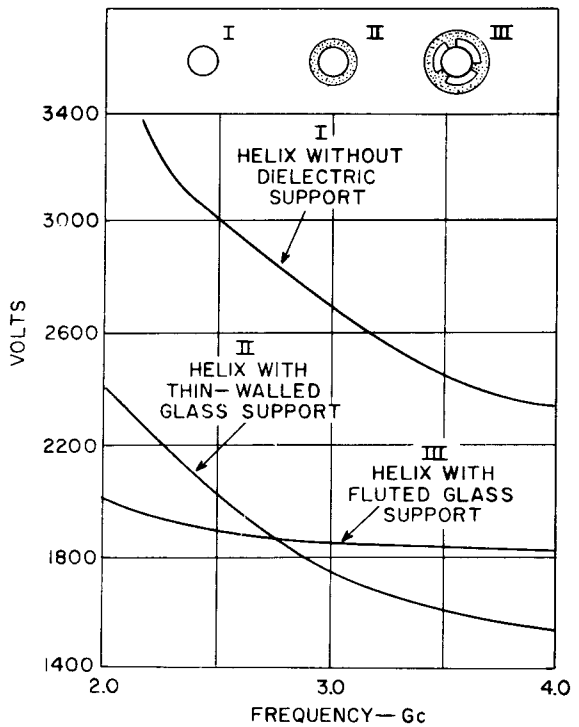


Figure 12. Effect of the Dielectric Loading on Helix Dispersion Resulting from Different Types of Support Structures

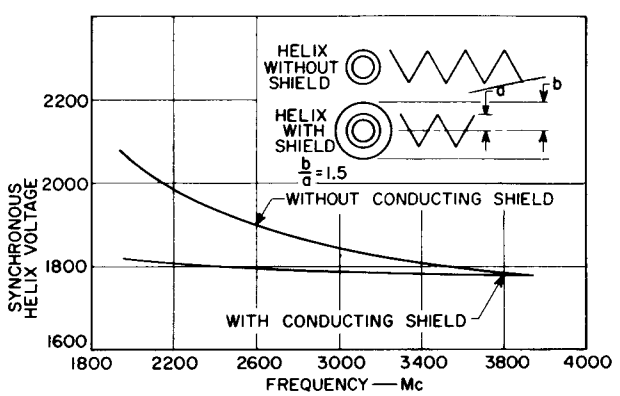


Figure 13. Curves Showing how the Use of a Conducting Shield over the Helix Reduces Dispersion

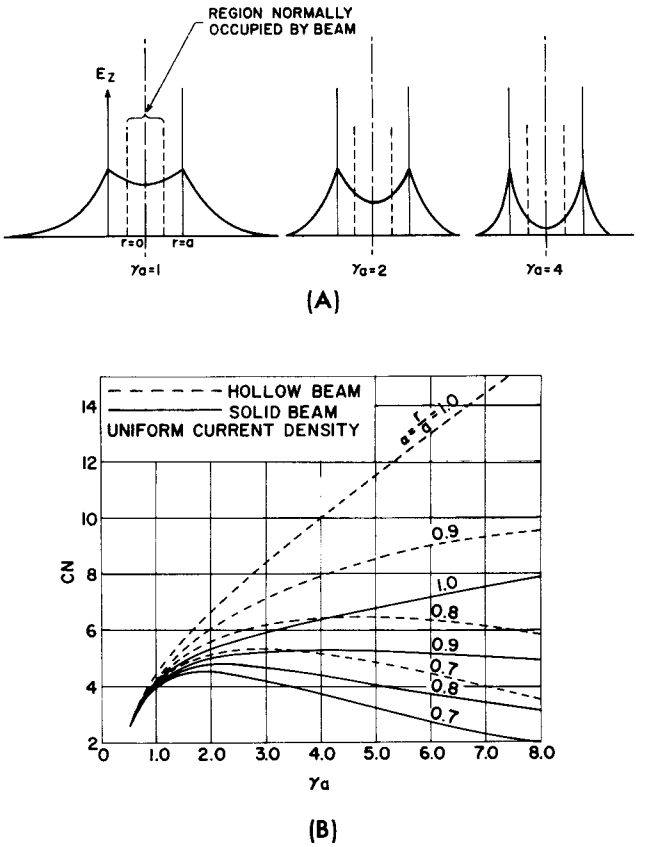


Figure 14. Effect of Beam Size and γa on Beam Impedance: (A) Showing that the amplitude of the axial component of the electric field on the axis of the helix (E_z) varies as a function of γa; (B) CN as a function of γa for several solid and hollow beams of different radii. CN is very nearly proportional to gain; thus, the curves show the variation of gain with frequency.

At normal ratios of beam-to-helix diameters of about 0.5 to 0.7, the impedance falls off at the higher frequencies while the number of helix wavelengths N, becomes smaller at lower frequencies. A plot of CN, which is roughly proportional to the gain [See Eq. (15)], as a function of γ_a for several beam diameters is shown in Fig. 14B.

If the tube is operated at a frequency at which γ_a is less than 1.0, the beam-field interaction is very efficient (high C), but the helix becomes dispersive and the beam and wave fall out of synchronism over a relatively narrow frequency range. For this reason, only narrow-band operation is obtainable at low values of γ_a.

A helix designed for wideband operation is a compromise between low γ_a, where the helix is a high impedance circuit, but dispersive and high γ_a, where the helix is less dispersive, but has a low impedance. Experience indicates that for most applications, a value of γ_a of 1.5 at the center frequency is a good compromise between efficiency and broadband operation.

Fig. 14 shows that at a high value of γ_a, the amount of the electric field interacting with the beam is highly

dependent on the ratio of the beam diameter to the helix diameter. Thus, for the typical beam diameter shown, a much larger proportion of the electric field inside the helix is interacting with the beam when γa is equal to 1 than when γa is equal to 4. However, if the beam diameter is increased, the impedance at high values of γa increases rapidly. This effect is shown quantitatively in Fig. 16 for solid and hollow beams. When this factor is included, the impedance becomes:

$$K = \frac{\left[F \gamma a\right]^3 \left[IRF\right]^3 \frac{c}{v}}{2} \qquad (24)$$

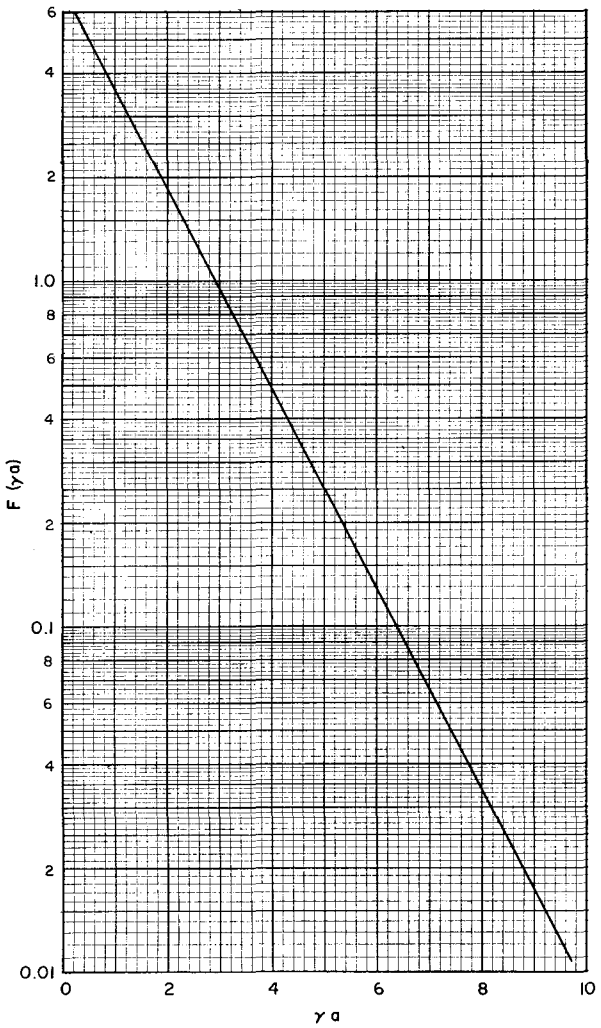


Figure 15. The Impedance Parameter $F_{(\gamma a)}$ as a Function of γa. The impedance on the axis of the helix is obtained from $F_{(\gamma a)}$.

The final correction factor to be introduced is Tien's impedance factor F, which describes the reduction in the electric field inside the helix as a result of the dielectric loading factor. In addition, the curve for $F_{(\gamma a)}$ was derived for a sheath helix and Tien's factor includes the impedance reduction resulting from the use of tape or wire instead of a conducting sheath. This factor is plotted in Fig. 17 as a function of ka for varying values of DLF. Thus, the final expression for K is written as follows:

$$K = \frac{\left[F \gamma a\right]^3 \left[IRF\right]^3 F \frac{c}{v}}{2} \qquad (25)$$

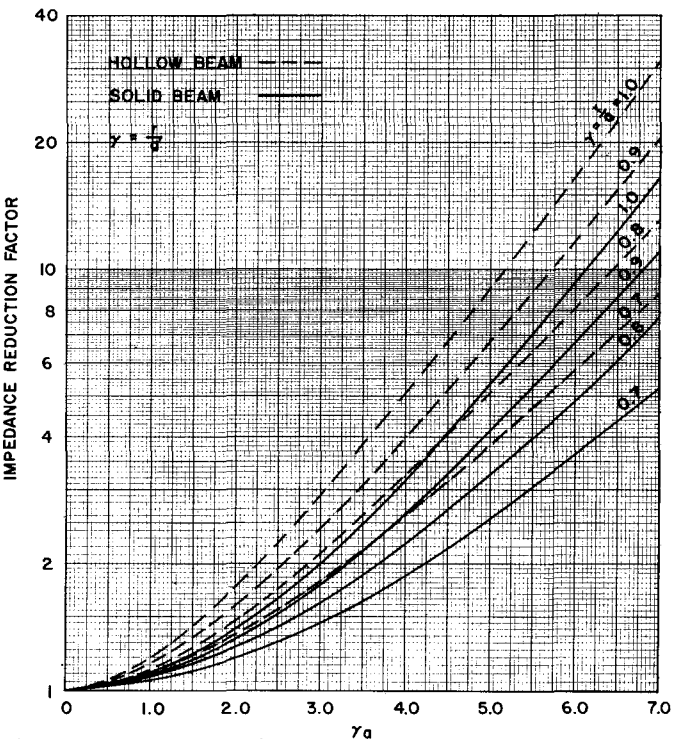


Figure 16. Family of Curves Relating the Impedance-Reduction Factor to γa for Several Hollow and Solid Beams with Different Radii. This factor is used to modify the impedance on the axis for solid or hollow electron beams.

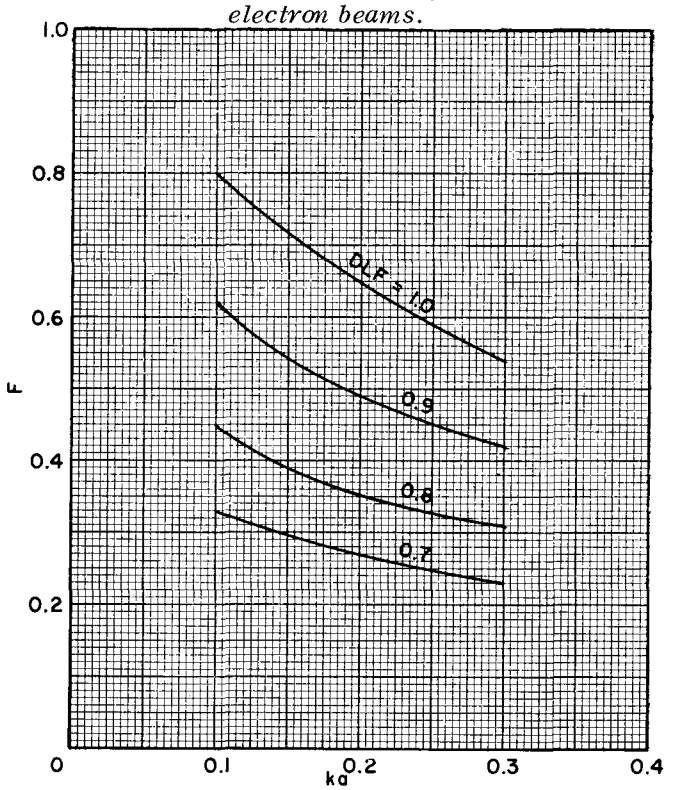


Figure 17. Curves Showing the Impedance-Reduction Factor F as a Function of ka for Several Values of the Dielectric Loading Factor

Limitation of the Helix

The helix, although simple to make and having excellent rf characteristics, is normally limited in use to the lower power level traveling-wave tubes for several reasons.

First, the helix has rather poor power-handling characteristics. The structure is inherently light and skimpy; at higher frequencies it is small and difficult to cool. This factor primarily limits the cw capabilities of a helix-type traveling-wave tube, as shown in Fig. 18. The upper regions of this range are reached only by providing some sort of liquid cooling to the dielectric supporting the helix or to the helix itself. However, as efforts in this direction are increased, dielectric loading and the complexity of the structure increase to the point where other types of circuits are preferable.

Second, the use of helices for high-power traveling-wave tubes is limited by the interference of other space harmonics. Two effects are felt: first, a reduction of impedance in the desired space harmonic and, second, the occurrence of backward-wave oscillations at a frequency near the operating frequency of the tube.

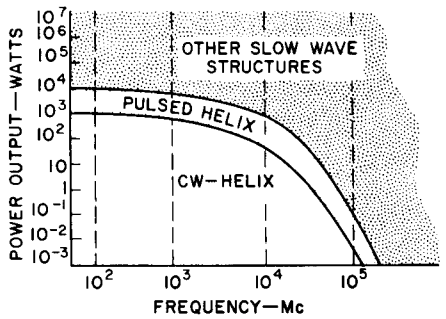


Figure 18. Power Limitations on Traveling-Wave Tubes Using the Helix Type of Slow-Wave Circuit

As described earlier, γ_a for a traveling-wave tube must be between 1.0 and 1.6 to maintain both efficiency and bandwidth. Thus, as the voltage is increased to obtain higher power levels, γ_a remains fixed and ka, which is $\gamma_a \frac{v}{c}$ increases. Fig. 17 shows that Tien's factor decreases with increasing ka, thus decreasing the impedance at higher helix voltages. Fig. 19 is a plot of helix voltage as a function of ka for both the forward- and backward-wave space harmonics of a helix. At the point where ka is 0.5, the phase velocity of both space harmonics is the same. The tube can act as a backward-wave oscillator at the frequency defined by a ka of 0.5 while operating as a forward-wave amplifier at some lower frequency. If the normal operating range of the tube is as high as ka equal to 0.4, then it may be difficult to suppress the backward-wave oscillation without affecting operation in the desired band. This condition normally occurs when the helix voltage is about 10,000 volts or greater or at lower voltages if the current is high or the tube is long and thus limits the maximum peak power available from the helix type traveling-wave tube.

OTHER SLOW-WAVE STRUCTURES

Many structures have been devised which in whole or part circumvent the limitations of the helix at high-power levels. In general, these structures have a considerable amount of stored energy which cannot interact with the beam and a much narrower bandwidth than the helix. Typical of such structures is the folded waveguide which was shown in Fig. 4. In this structure, the wave follows the waveguide at a phase velocity determined by the frequency and the dimensions of the waveguide, while the electron beam, following a more direct path, travels at a velocity determined by the beam voltage. By properly arranging the length of the path in the waveguide, the arrival of the wave at each point where the beam crosses the waveguide can be made to coincide with the same electrons which the wave met at the previous crossing point. In this way, the beam and wave are in synchronism down the length of the structure and interaction takes place.

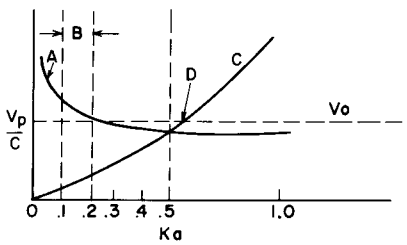


Figure 19. Phase Velocity v_p/c as a Function of ka for Both the Forward and the Backward Space Harmonic of the Wave on the Helix

A single cell of a practical structure is pictured in Fig. 20. The folds of the waveguide are brought together to a common wall and coupling from cell to cell is accomplished through the coupling slot. This is a very rugged and massive structure compared to the helix and is able to handle considerably higher average-power levels. The useful space harmonics of this structure are at high values of ka. This permits operation at voltages well above the 10-kilovolt limit of helices.

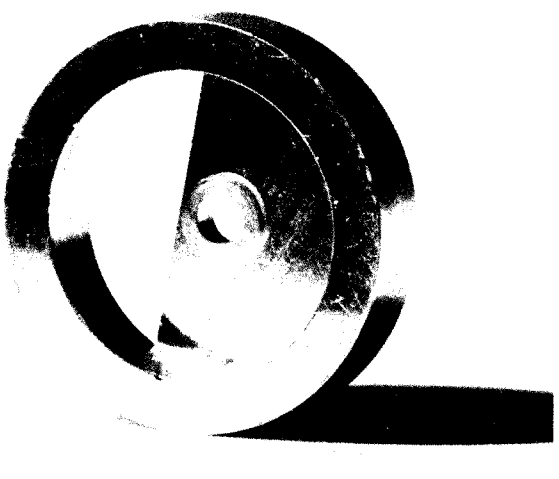


Figure 20. Single Cell of the Folded-Waveguide Type of Slow-Wave Structure

An excellent survey of many other types of slow-wave structures is given in an article by Nalos[3] which describes some of the types that are in use and the features that are desirable for specific applications. A

more detailed description of the characteristics of some slow-wave structures is given by Belohoubek.[4] The reader is referred to these and other references for the actual choosing and design of structures for high-powered traveling-wave tubes. However, as an introduction to these papers, a brief description of how these slow-wave structures are measured and compared is given below.

The characteristics of these structures is usually described by a Brillouin diagram, which relates the phase shift per cell to the frequency. A typical Brillouin diagram for the folded waveguide is shown in Fig. 21. The phase shift per cell is the amount of phase shift in the rf wave at a given instant in the distance from one interaction region to the next. This phase shift is usually given in terms of βL, where $\beta = \omega/v_p$ and L is the period of the structure. βL is equal to 2π or a 360-degree phase shift for that value of frequency at which the guide wavelength λ_g is equal to the period of the structure.

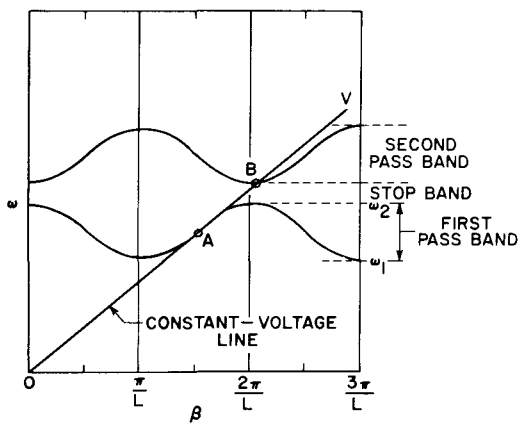


Figure 21. $\omega - \beta$ Diagram for the Folded-Waveguide Slow-Wave Structure

$$\beta L = \frac{\omega}{v_p} L = \frac{\frac{2\pi C}{\lambda_o}}{v_p} L = \frac{2\pi}{\lambda_o \frac{v_p}{C}} L = \frac{2\pi}{\lambda_g} L \qquad (26)$$

For values of βL less than 2π, less than one rf cycle is completed in passing from one cell to the next. For values of βL greater than 2π, more than one rf cycle is completed.

Notice that

$$\frac{\omega}{\beta L} = \frac{\omega}{\frac{\omega}{v_p} L} = v_p \qquad (27)$$

Thus, the slope of a line drawn from the origin to a point on the curve is proportional to the axial-phase velocity at that point. The slope of the curve itself is related to the group velocity of the wave which indicates the direction of power flow in the structure. Therefore, in the region where $\pi < \beta L < 2\pi$ the phase and group velocity are nearly equal and in the same direction. In addition, the phase velocity is nearly constant over the available frequency band from ω_1 to ω_2 and indicates little dispersion and, consequently, good forward-wave interaction over most of the available bandwidth. The constant voltage line intersects the second pass band at point B, however, which would indicate interference from possible oscillations at that frequency.

Also of interest is the region where $0 < \beta L < \pi$ in which the phase velocity is still positive, but the group velocity is still positive, but the group velocity is negative, indicating power flow in a direction opposite to that of the phase velocity. Note that for a given voltage there is only one frequency at which interaction can take place; thus, this is a backward-wave device, voltage-tunable over the pass band.

The determination of the $\omega - \beta$ diagram is important in the design and evaluation of slow-wave structures because it gives information on the bandwidth, voltage, and possible interference from other modes. The technique used is to take a short length of the slow-wave structure, usually about 5 cells, short the ends of the structure and determine the resonant frequencies. These resonant frequencies can then be related to certain phase shifts per cell and plotted to form the $\omega - \beta$ diagram.

The impedance of such structures can be obtained from perturbation of the fields on the axis of the shorted structure by means of small dielectric beads. The impedance can then be calculated from measurements of the frequency change due to the introduction of the dielectric. A detailed description of these measurements is too complicated to be included here, but several good descriptions of techniques are available in literature.[5]

Figs. 22 and 23 show, respectively, a cross-section drawing of a structure designed for 1-kilowatt peak power over a 20 per cent bandwidth at S-band and the $\omega - \beta$ diagram plotted for the same structure. This structure may be scaled to other frequencies, with the voltage constant, by scaling all dimensions to the change in wavelength. The voltage is changed by scaling the period only by the square root of the voltage. Variations of dimension "w" will vary the bandwidth. A larger coupling slot will increase the bandwidth but decrease the impedance.

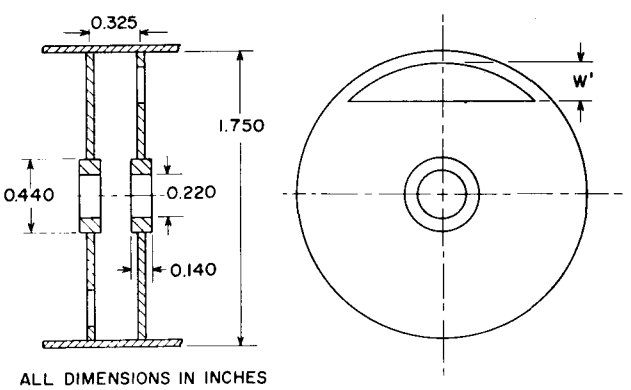


Figure 22. Cross-Sectional Drawings of a Folded-Waveguide Slow-Wave Structure Designed for S-Band Operation at 1 Kilowatt of Peak Power over a 20 Per Cent Bandwidth

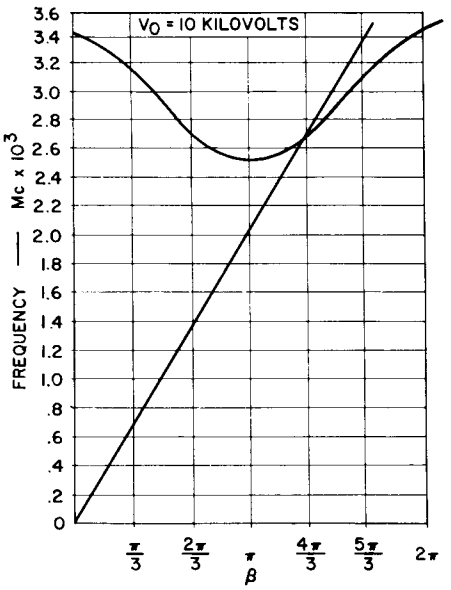


Figure 23. $\omega - \beta$ Diagram for the Folded-Waveguide Structure Shown in Fig. 22

ELECTRON GUN

The design of the electron gun can be extremely critical in the ultimate performance of a traveling-wave tube. The formation of a well-defined beam of correct diameter is important in obtaining the proper rf characteristics of a tube. A gun producing a beam smaller than required for a given helix or slow-wave structure lowers the interaction impedance, particularly at the high-frequency end of the band. Conversely, a beam too large or not well defined is difficult to focus. In high-power tubes in which interception on the slow-wave structure can be sufficient to cause severe thermal problems, a well-defined beam of the proper diameter is vital to the successful operation of the tube.

Many different types of electron guns are in use on traveling-wave tubes. One of the most common types is the convergent-flow gun shown in Fig. 24. A related gun, the parallel-flow type, is shown in Fig. 25. Fig. 26 is a hollow-beam parallel-flow gun. Fig. 27 is a convergent-flow hollow-beam gun. Fig. 28 is a magnetron injection gun. A typical low-noise gun is shown in Fig. 29. Each of these types has a specific application. The most commonly used is the convergent-flow solid-beam electron gun. Because of the widespread use of this particular gun, the following description is fairly detailed, and some design procedures are also discussed. For the other types of guns, only the operation and literature references for design are given.

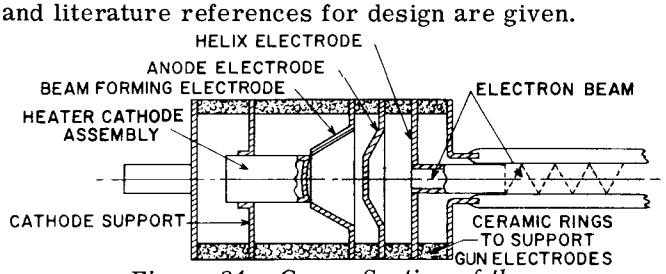


Figure 24. Cross Section of the Convergent-Flow Electron Gun

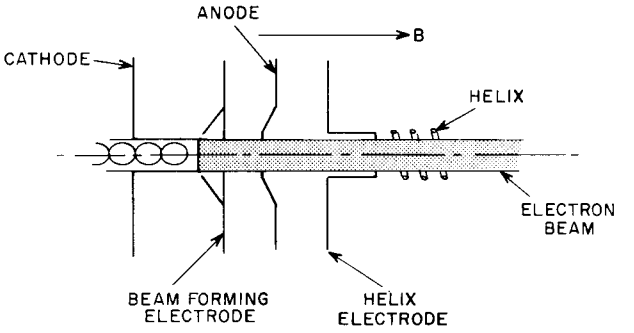


Figure 25. Schematic Diagram of a Parallel-Flow Electron Gun

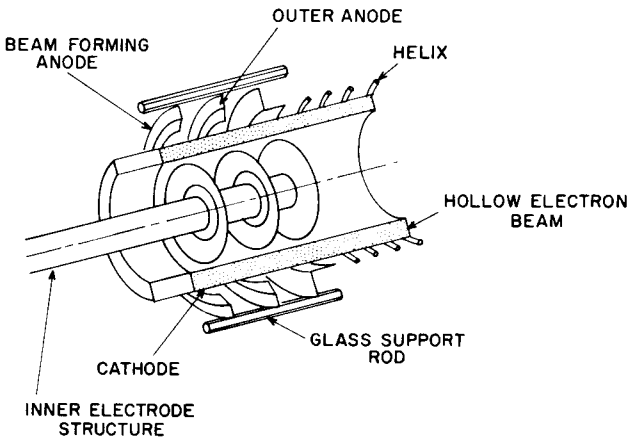


Figure 26. A Hollow-Beam Type of Parallel-Flow Electron Gun

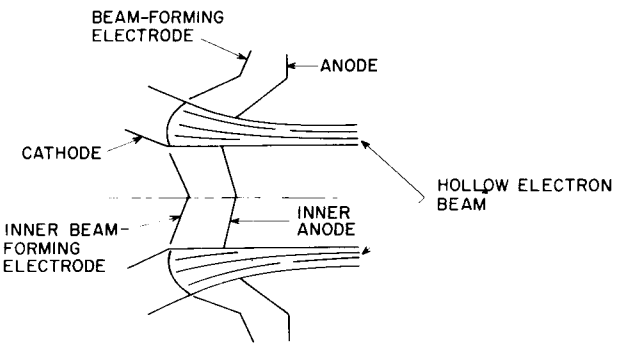


Figure 27. A Hollow-Beam Type of Convergent-Flow Electron Gun

Convergent-Flow Electron Gun

The use of a convergent-flow, solid-beam gun is desirable for several reasons. A higher current density in the beam can be obtained than is normally available from a cathode surface by converging the beam in the gun region. Convergence ratios of 2:1 to 10:1 are readily available and provide extremely high beam-current densities in the interaction region of the tube. An initially convergent beam is necessary for several of the focusing schemes, described later, in which the gun region must be shielded from the magnetic field and the beam must enter the magnetic field with parallel elec-

tron trajectories. From the parallel-flow gun the beam would normally be divergent as it leaves the anode of the gun. However, with a convergent-flow gun, a beam minimum occurs some distance from the anode, at which point the electron trajectories are parallel, the radial velocities of the beam are at a minimum, and the magnetic field can be introduced.

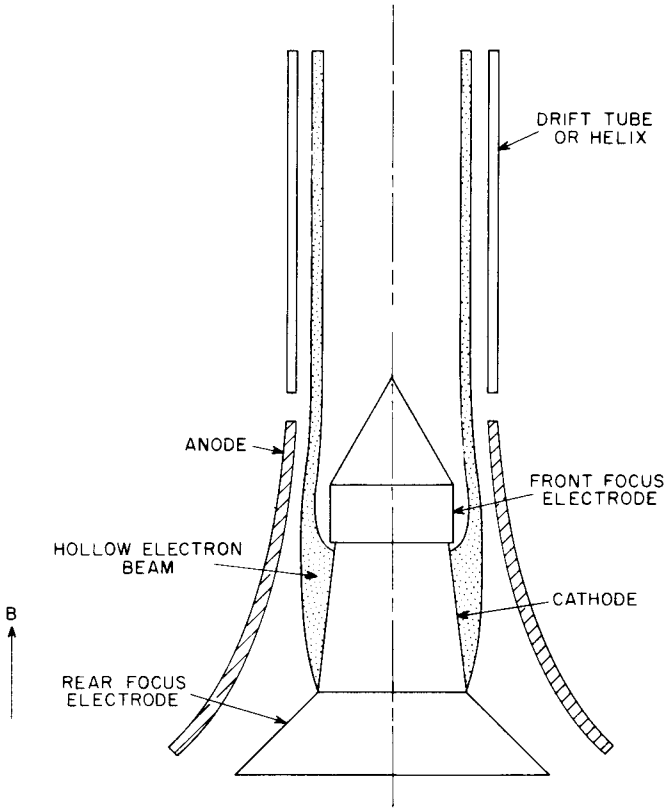


Figure 28. Cross Section of a Magnetron-Injection Electron Gun

Considerable analytical background is available which makes it possible to design a convergent-flow electron gun having a microperveance of less than one, where perveance p is equal to $I/V^{3/2}$. A particularly good reference by Danielson, Rosenfeld, and Saloom[6] is found in the Bell System Technical Journal. The general approach to the analytical method is as follows: It is assumed that in two concentric spheres, the inner surface of the outer sphere emits electrons to the inner sphere. By the law of minimum astonishment, the electrons all travel in a straight line from the outer sphere to the inner sphere, with none of their paths crossing. The object in designing an electron gun is to cut out a segment of these spheres and form a potential along the boundary of the beam so that the cut out portion of the beam is not affected by the missing portion of the spheres. Presumably, the electrons travel in smooth laminar paths from the cathode to an anode. Reference 6 provides the correction required for the hole in the anode and also includes the effect of the thermal velocities of the electrons.

For a perveance greater than one, the design curves break down because the hole in the anode through which the beam must pass becomes quite large compared to the spacing between the anode and the cathode. Two methods are available when designing a gun having a perveance greater than one. The first method involves setting up the boundary conditions for the beam in an electrolytic tank and experimentally determining the shape of the electrodes which produce the proper electric field along the edge of the beam. The second is to scale from an existing gun. The first of these methods has been adequately described in the literature.[7,8]

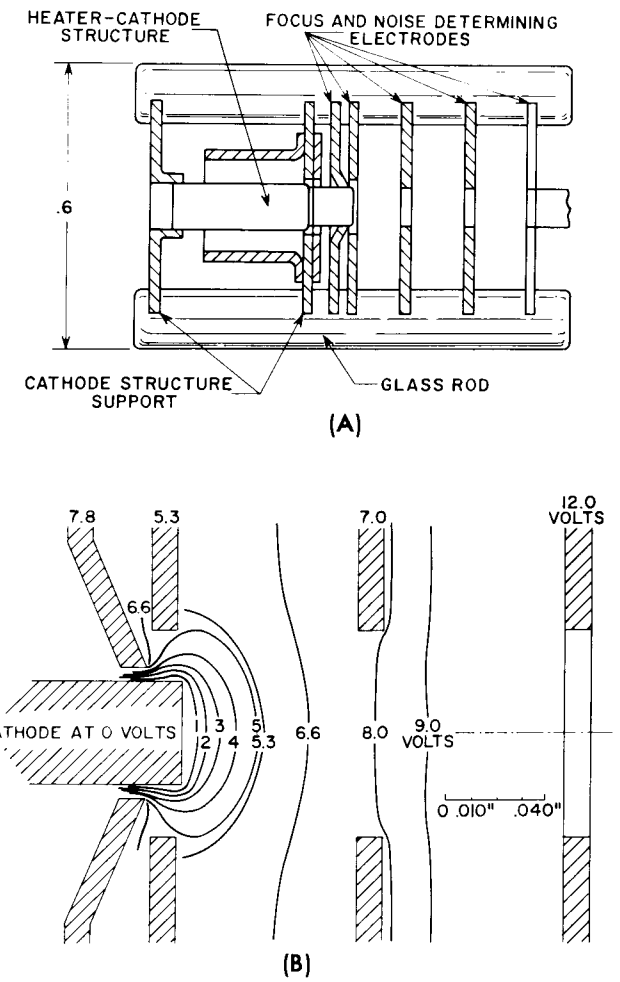


Figure 29. A Typical Low-Noise Electron Gun: (A) Cross section of the gun structure; (B) Electrode arrangement and distribution of the equipotential lines in the gun

An electron gun, the dimensions of which were empirically determined by Siekanowitz in 1954 and which has been modified to produce beams with a perveance ranging from 0.5 to 1.7 at a convergence ratio of 4 to 1, does provide an excellent starting point for scaling of a gun design. The critical dimensions of this gun are shown in Fig. 30; the minimum beam diameter has been normalized to unity. This gun, operated at a perveance of 1.25, produces a minimum beam diameter of 0.110, from a cathode diameter of 0.400 with the beam minimum located 0.550 inch from the anode aperture. The original and scaled models of the gun have been used in several tubes and have provided excellent focusing at very nearly the minimum theoretical value of the mag-

netic field. Figs. 30B, C, and D, show how the various parameters of this gun were modfied in order to vary the perveance from 0.5 to 1.75. In each case, the cathode diameter remained the same and the beam-minimum diameter remained the same within about 10 per cent.

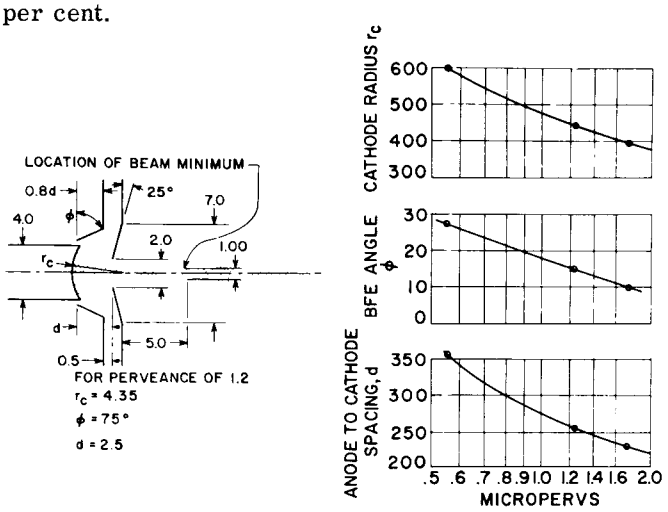


Figure 30. Schematic Diagram Showing the Electrode Dimensions and Spacings of a Modified Siekanowitz Electron Gun. Graphical data show how the parameters of this gun may be modified to vary the perveance over the range of 0.5 to 1.75 micropervs.

The performance of this gun has been consistently excellent. For example, in one tube, a 100 milliampere beam current at a beam voltage of 2000 volts can be focused through a 0.175-inch I.D. helix with only 1 per cent interception at a peak magnetic field of 350 gausses in a periodic permanent-magnet structure. In another case, a 50-milliampere beam at 1100 volts was focused with 3 per cent interception through a helix with an I.D. of 0.130 inch at a peak magnetic field of 450 gausses. In both cases, the peak magnetic field is very near the minimum theoretical value. Such good focusing can only be obtained with a beam generated from a gun having few aberrations.

Parallel-Flow Gun

When the magnetic field is permitted to thread through the cathode itself and the current density in the beam is not too great, a parallel-flow gun may be used. The derivation of the analytical approach for this gun is similar to that of the convergent-flow electron gun except that instead of two concentric spheres there are two infinite planes, one of which emits electrons to the other. In this case, the electric field along the beam in the gun region is a simple function of the distance between the planes and can be readily duplicated with electrodes by following the procedure of Pierce.[7] However, because of the hole in the anode, the beam diverges as it leaves the gun and, therefore, this gun is practical only in cases where the magnetic field threads through the gun region. The disadvantages of this type of focusing will be explained later.

Hollow-Beam Parallel-Flow Gun

The hollow-beam parallel-flow gun is essentially similar to the parallel-flow gun, except that instead of producing a solid circular beam, a long thin ribbon beam is formed which is folded back upon itself to form a hollow beam. This type of gun is extremely useful in cases where the electron beam must be very near the interaction structure for good performance. As discussed earlier, it is particularly useful to maintain high impedance when operating at high values of γ_a.

In this type of gun, extremely high perveances may be obtained. For instance, if the circular cathode is spread out in a straight line and then divided up into squares, the perveance of each square is limited to about one, as described above for the convergent-flow gun. However, with a sufficiently large length-to-width ratio, there may be 10 to 100 such squares. Therefore, the total perveance of the gun as a whole can be 10 to 100 times the perveance of a single square; it is not unusual to have a perveance of 10 to 100.

Convergent-Flow Hollow-Beam Gun

The convergent-flow hollow-beam gun has been rarely used in experimental tubes. As shown in Fig. 27, the inherent difficulty of obtaining a well-defined beam from a hollow convergent cathode and properly directing the beam down the axis of the tube is so formidable that no clear-cut analytical method has been found to describe it. The experimental results to date, which have not been very encouraging, are included in the references.[9, 10]

Magnetron-Injection Gun

The magnetron-injection type of gun has received increasing attention in recent years because of its extremely high convergence ratio and the fact that the beam can be readily focused in a uniform magnetic field. As shown in Fig. 30, the electrons are drawn away from the cathode by a slight positive potential on G_1, but are constrained to move in an axial direction because of the magnetic field. In this way, a rather well-defined hollow beam of very high current density (compared to the current density of the cathode) can be formed. Because no flux actually passes through the cathode surface, the electrons leaving the gun have a rotational velocity which aids focusing. The design procedure for this type of gun is given in reference 11.

Low-Noise Gun

Special design problems exist in the development of electron guns for low-noise traveling-wave tubes because, in these tubes, the gun serves a dual function: not only is the electron beam generated and defined by the gun, but the gun must also maintain the proper beam-entrance conditions for minimum noise at the point of entry into the rf interaction region.

Modern theory considers two essentially uncorrelated sources of noise: velocity fluctuations and current fluctuations of the electrons leaving the cathode. The beam can be considered as a propagating medium for these noise waves and an analogy has been constructed between the beam viewed as a space-charge-

wave transmission medium and the familiar transmission line. The gun may be considered to act as a transducer which must be carefully designed to match the impedance (ratio of voltage or velocity fluctuations to current fluctuations) at the cathode plane to the desired impedance at the helix. Several of the familiar transmission-line transformers can be used as models. The exponential transformer comes closest to fulfilling the requirements of the low-noise gun. The RCA "three-region gun" is a transformer of this type.

A discussion of design of low-noise guns can be found in a thorough study by R. W. Peter in reference 12. The theory of low-noise guns is presently confused and the design parameters in the region of the cathode are now undergoing constant revision. At the time of publication, no survey article has appeared which covers the developments of the past few years since Peter's article was written.

FOCUSING

Many schemes have been devised for focusing the beam in a traveling-wave tube. Most efforts have been directed at reducing the size and the weight of the focusing structure. Several of the better-known types of focusing which have actually been put into use in production or developmental tubes are described below. In most cases, the more esoteric forms of focusing are some modification or combination of these focusing methods.

Magnetic Focusing

Magnetic focusing was the first method used to confine the beam in the interaction region of traveling-wave tubes. After fifteen years, magnetic focusing is still used in over 95 per cent of traveling-wave tubes. During that time, the techniques and the means of supplying the field have become increasingly refined.

Basically, there are two types of magnetic focusing. The first and most commonly used in the earlier days of the industry is magnetically confined flow, otherwise known as the "brute force" method. The second and more commonly used method today is magnetically focused flow, frequently called Brillouin flow.

Confined Flow. When a strong magnetic field is directed along the axis of the tube so that the flux is threaded through both the cathode and the interaction region, the magnetic field constrains the electrons to follow a relatively straight path from the cathode down the length of the tube. If the beam is not well focused, the magnetic field is merely increased until it reaches a value sufficient to provide the desired beam transmission.

Radial forces on the electrons come from the space-charge forces generated by the mutual repulsion of the electrons; thus, the higher the current density in the beam, the stronger the magnetic field required. The path of an individual electron is shown in Fig. 31. Space-charge forces produce a small radial component in the electron velocity which, under the action of the magnetic field, results in a circular motion of the electron when viewed from the end. In traveling down the tube, the electron follows a helical path. The strength of the magnetic field determines the diameter of the circle described by the electrons and thus determines the definition of the beam. For beams of moderate current density such as in a ten-watt S-band tube, about 500 to 600 gausses is adequate for focusing.

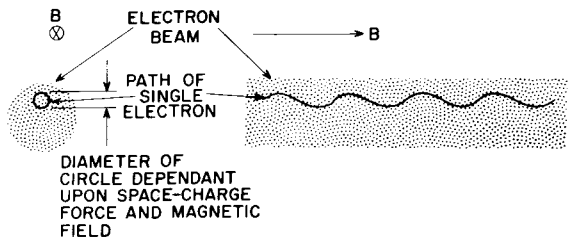


Figure 31. Path Traced by an Individual Electron in Confined Magnetic Flow

A parallel-flow gun must be used with this type of focusing because the electrons are under the influence of the magnetic field from the time they leave the cathode surface. Use of any of the convergent-type guns would involve cutting lines of flux during the convergence of the beam and, consequently, destroying the action of the gun. In the parallel-flow gun, both the action of the electrostatic lens in the gun and the magnetic field tend to keep the electrons moving in a straight line.

With the exception of its use in low-noise tubes where the magnetic field requirements are determined more by the noise-figure properties than by the density of the beam, this type of focusing is not generally in widespread use because it requires a stronger magnetic field than focused flow.

Focused Flow. Focused flow is considerably more sophisticated and requires less magnetic field and a smaller size magnet than confined flow. The most common type of focused flow, known as Brillouin flow, in which two counterbalancing forces on the electrons in the beam tend to keep the beam on the desired path. The electron beam is generated, usually by a convergent-flow gun, and the magnetic field starts abruptly at the point at which the beam reaches its minimum diameter and the radial velocities are at a minimum. Bush's theorem states that upon entering the magnetic field the electrons begin to rotate about the axis with a rotational velocity which is proportional to the distance of each electron from the axis of the beam and the strength of the magnetic field. Thus, we have set up the conditions shown in Fig. 32; an electron beam travels through the interaction region with each electron following a helical path about the axis of the tube at a constant radius.

Under the ideal conditions, three forces act on each electron. First, there are the normal space-charge forces which tend to increase the radius of the electron path. Second, there is the centrifugal force, outward, of the spiraling electrons which also tends to increase the radius. Third, the spiraling electrons are cutting lines of flux and experience an inward Lorentz force. When the conditions for Brillouin flow have been satis-

fied, these forces balance out to zero and the electrons are focused down the length of the tube. The conditions for Brillouin flow B_0 are shown by the following equation:

$$B_0 = \frac{326\, I^{1/2}}{V^{1/4} a} \text{ gauss} \qquad (28)$$

where I is the beam current in amperes, V is the beam voltage in volts, and a is the radius of the beam in inches. In practice, the actual field required is frequently about 1.3 times the value obtained from this equation.

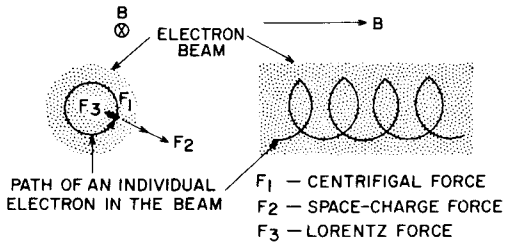


Figure 32. Path Traced by an Individual Electron in Focused Magnetic Flow

If an electron is perturbed slightly, either by entering the magnetic field with an initially slight radial velocity or as a result of some perturbation during transit, there are forces which tend to restore the electron to its original orbit. A force diagram of the relative forces on a given electron having an initial radius r_0 is shown in Fig. 33. As the orbital diameter of the electron becomes larger, greater force tending to decrease the diameter is created because more flux lines are cut. At the same time, the centrifugal and the space-charge forces decrease. On the other hand, if the diameter of the electron path becomes smaller, the space-charge force and the centrifugal force increase and the magnetic force decreases, tending to push the electron out to its original path. Thus, a potential valley is formed with a minimum at the equilibrium radius. The steepness of the walls of this potential valley determine the stiffness of the focusing. Actually, if electrons are displaced at any time, an oscillatory motion takes place in the diameter of the beam as a whole, and results in beam scalloping about the initial radius r_0.

This type of focusing requires less field than confined flow, but also requires a certain optimum field. In other words, unlike confined flow in which the magnetic field is merely increased to improve the beam definition and transmission, increasing the magnetic field results in a severe inward scalloping of the beam and a reduction in the average beam diameter which can be detrimental to the rf performance of the tube.

Sources of Magnetic Field. Generating the magnetic field for focusing traveling-wave tubes has been considerably advanced in recent years. Initially, the solenoid was used to generate a uniform magnetic field because performance was easily predicted and the magnetic field could be readily varied to find the optimum field for best focusing. However, as traveling-wave tube requirements become more stringent, especially in airborne applications where power consumption and weight were extremely important, solenoids were replaced with various types of permanent magnets.

Solenoids. Solenoids are still frequently used today in special applications such as low-noise tubes in which the magnetic field must be carefully controlled and transverse fields held to a minimum, and in tubes for high-temperature service, in which the commonly-used magnetic materials cannot be used. Improvements in the insulating materials for magnet wires will eventually permit the use of solenoids in ambient temperatures up to several hundred degrees centigrade.

Permanent Magnets. Permanent magnets, although weighing as much or more than a well-designed solenoid, have the advantage of requiring no power to maintain the magnetic field. Although largely superseded by periodic magnet focusing, which requires considerably less magnetic material, the straight-field permanent magnets still are useful in focusing very-low-noise traveling-wave tubes and tubes requiring a very high magnetic field for a relatively short distance. In the latter case, the present material limitations prevent the use of periodic permanent magnetics. The design procedure for the straight-field permanent magnet has been described by Glass[13] and Schindler[14]. These magnets can be used for either confined flow with the flux threading the cathode or Brillouin flow with the gun suitably shielded from the magnetic field.

Periodic Permanent Magnets. Periodic permanent-magnet focusing was the first major breakthrough in making the traveling-wave tube a light-weight portable device. As mentioned under "Brillouin Flow," the centrifugal and space-charge forces on the electron are counterbalanced by the magnetic field. However, in the equation describing this relationship, the magnetic field term appears as B_0^2. Therefore, the sign or polarity of the magnetic field can be neglected. Thus, if the magnetic field changes polarity periodically down the length of the tube, focusing is unaffected.

The arrangement of a periodic permanent-magnet focusing stack is shown in Fig. 34. The magnetic material is in the form of rings of high-coercive-force material, usually a barium ferrite, and the shims are formed of magnetically soft iron, usually Armco iron. The ring magnets are alternately magnetized so that they oppose each other and form the indicated magnetic field along the axis.

In a periodic permanent magnet, the stray external fields tend to cancel each other and thus produce a much greater percentage of the usable field in the center of the magnet stack, where it can be efficiently used to focus the beam. In general, as the length over which a magnetic field is to be maintained is halved, the weight of the magnet required to maintain that field is reduced by a factor of 8. Thus, if a single permanent magnet is replaced by a periodic permanent magnet having some ten periods, there will be a theoretical reduction of weight of 100 to 1. While this is not normally obtained in practice because of the greater demagnetizing forces in a periodic permanent magnet, a normal re-

duction of greater than 10 to 1 in the weight of a focusing structure is readily obtainable.

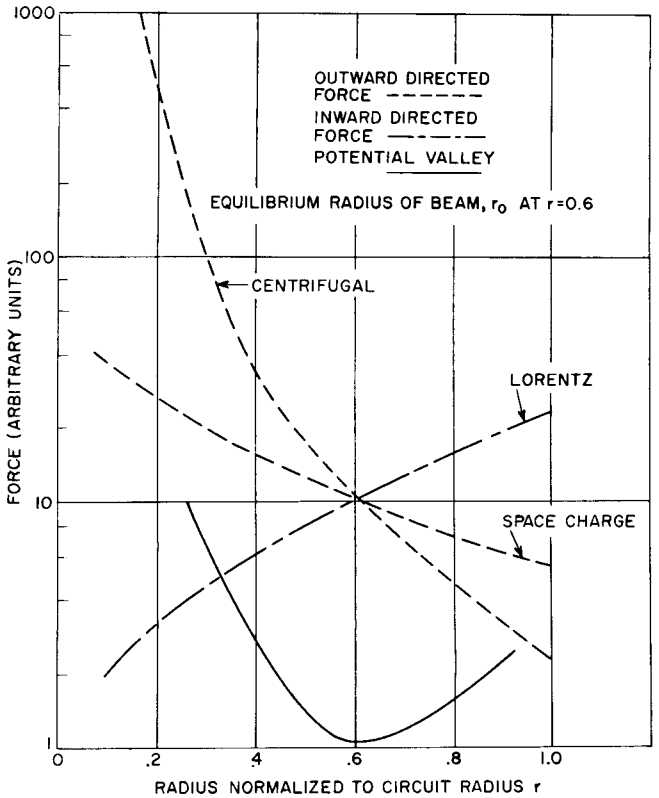


Figure 33. Diagram of the Forces Acting on a Single Electron in a Brillouin-Flow Beam

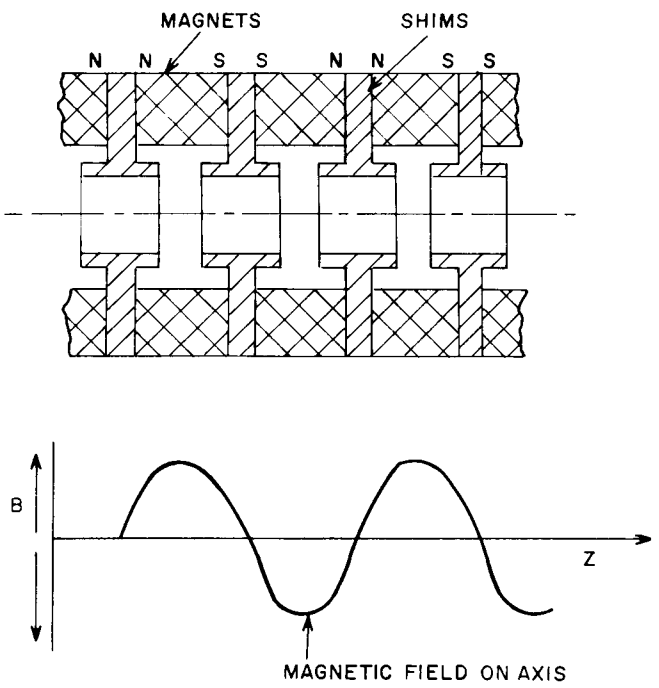


Figure 34. Cross-Sectional Diagram of a Typical Periodic-Permanent-Magnet Focusing Stack and the Distribution of the Magnetic Field along its Axis

Many design procedures have been devised to calculate the magnet dimensions required to provide the given magnetic field on the axis for a given period. The most recent, by Schindler,[15] includes probably the most accurate and certainly the fastest available designing for these magnets.

Two conditions must be met to obtain satisfactory focusing with periodic permanent magnets: the proper rms magnetic field must be obtained and the periodicity of the focusing structure must be within certain limits. Because the magnetic field along the axis is sinusoidal, the Brillouin field B_0 must be multiplied by a factor of 1.4 to obtain the peak magnetic field B_m. The proper magnetic field results in the same type of focusing discussed under Focused Flow.

The periodic nature of the magnetic field imposes a limit on the allowable period of the magnet structure. The rising and falling of the magnetic field along the axis of the beam causes a scalloping of the beam in which the scallops have the same period as that of the magnet stack. As in Brillouin flow, a focused electron beam also has another scallop having a periodicity related to the restoring forces on the electrons. If the scallops from the periodicity of the magnetic field and the scallops from the restoring forces have approximately the same frequency, the beam tends to expand rapidly. The condition which sets the allowable limits of the period of the magnet structure is expressed by the following equation:

$$L = C \frac{V^{1/2}}{B_m} \qquad (29)$$

where L is the period in inches, V is the beam voltage in volts, B_m is the peak magnetic field in gausses, and C is a constant between 0 and 16. If C is maintained below 16, the scallops are not amplified and the beam is stable. In actual practice, the value of C ranges from 8 to 10. Above 10, the scalloping is excessive, and below 8, the demagnetizing forces encountered in the periodic stack are so high that the required magnet stack is difficult to construct or too great in weight.

Fig. 35 shows two typical helix-type traveling-wave tubes focused with periodic permanent magnet stacks. For the tube using helical couplers in which the input and output lines are of relatively small diameter coaxial cable, the line can be brought up directly through a hole drilled in either the magnet or the shim. In general, cutting through the magnet is best, because it causes less interference with the magnet circuit and is less likely to introduce transverse fields on the axis. In the case of the waveguide coupler, the waveguide is made as narrow as possible and is placed between two magnets. This interruption of the magnetic field does not seem to affect the focusing appreciably.

Fig. 36 shows how a periodic permanent-magnet structure is built into a folded-waveguide slow-wave structure. The walls of the cavity and the drift tubes are made of soft magnetic material, and the outside vacuum-envelope ring is made of copper. The soft iron extends out beyond the interaction space and the mag-

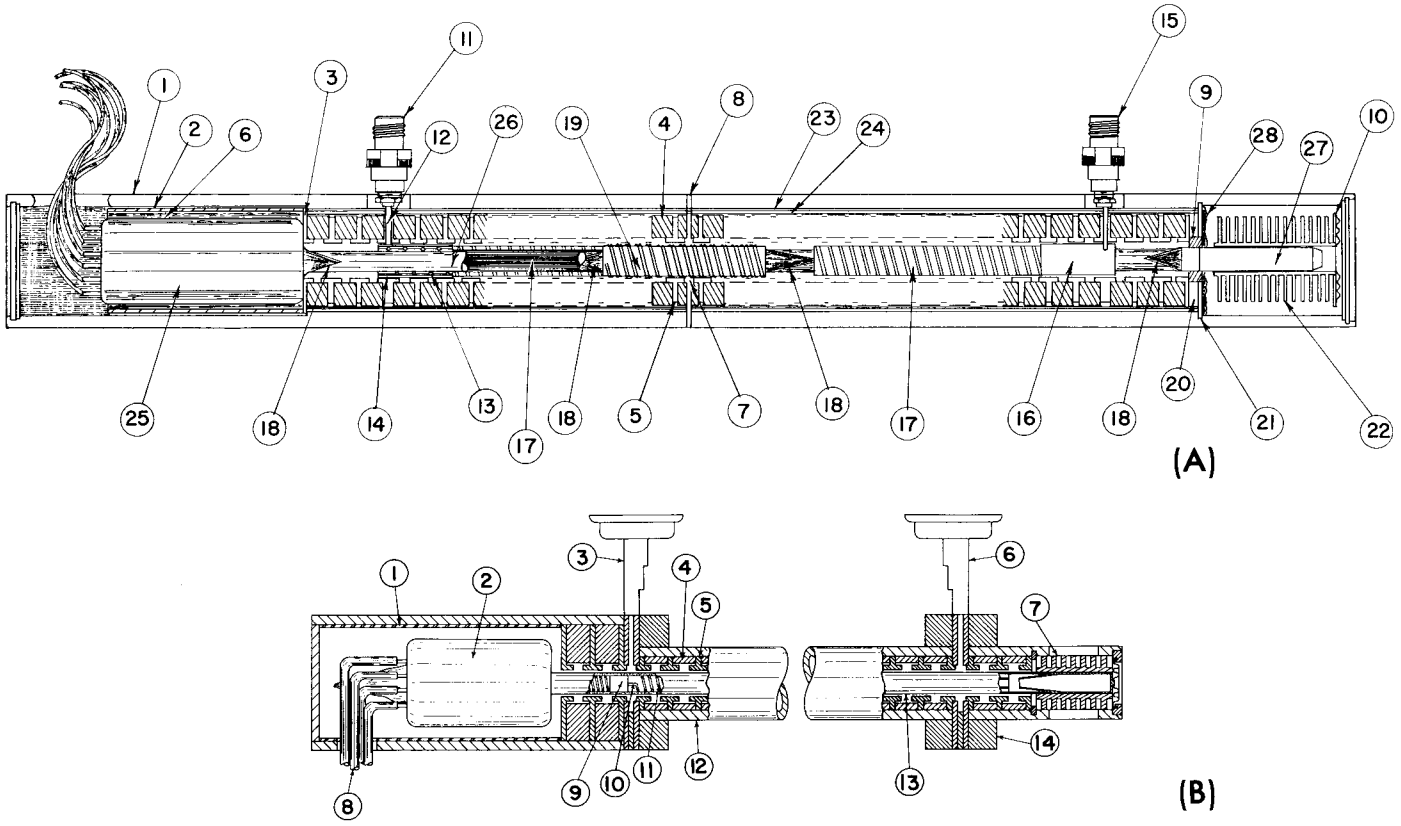


Figure 35. Cross-Sectional Drawings Showing the Arrangement of a PPM Stack in a Helix-Type Traveling-Wave Tube: (A) Tube using helical couplers; (B) Tube using waveguide couplers

nets, which are split for ease of assembly and are then placed between the shims. This type of construction has made possible X-band folded-waveguide traveling-wave tubes of relatively small size and weight with peak powers of 15 to 500 kilowatts and average powers up to 1000 watts.

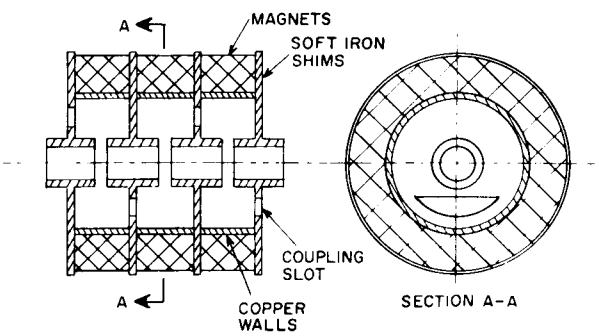


Figure 36. Arrangement of a PPM Stack Using a Folded-Waveguide Slow-Wave Structure

Electrostatic Focusing

Because of the need for small and light traveling-wave tubes for airborne equipment, missiles, and satellites, light-weight electrostatic focusing has been used to eliminate the magnet or solenoid entirely. Two major types of electrostatic focusing are used: (1) spacially alternating electric fields such as the Einzel Lens (Fig. 37) or (2) the radial electric field generated between two concentric cylinders. In addition, combinations of these two methods with magnetic fields can achieve greater stability of electron flow.

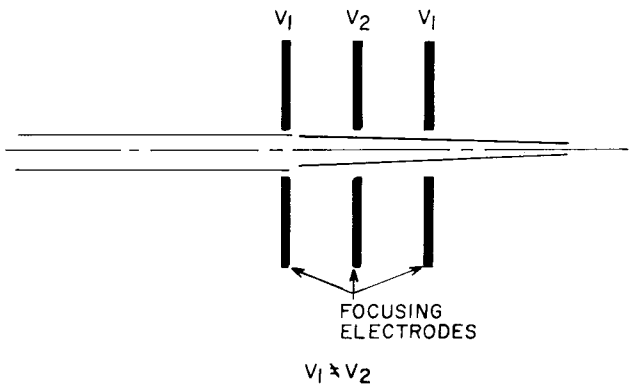


Figure 37. An Einzel-Lens Electrostatic Focusing Arrangement

Spacially Alternating Electrostatic Fields - Bifilar Helices. A bifilar helix (such as that shown in Fig. 38) having two different potentials applied to the two windings, has a net focusing force similar to that obtained with an Einzel lens. As an electron passes alternately high- and low-voltage helix wires, there is a net inwardly-directed force on the electron which is balanced

only by the space-charge forces of the electron beam. The force diagram, shown in Fig. 39, shows the relative amplitude of these forces on an electron placed at a distance r_0 from the axis in a helix with a mean radius of a. This form of focusing is not very rigid because no strong forces tend to keep an electron exactly in its original path. The inward-directing forces from the bifilar helix fall off very rapidly from the helix and the outward space-charge force is relatively weak. The forces in this diagram indicate considerably less stiffness than those of Fig. 33, which show Brillouin flow. Because the electrostatic force falls off so rapidly from the helix, it is generally believed that the beam becomes partially hollow as it passes through the helix and becomes essentially a "dirty" hollow beam. Because of the lack of rigidity, the maximum perveance which can be focused with a 3:1 ratio of the two helix voltages is about one.

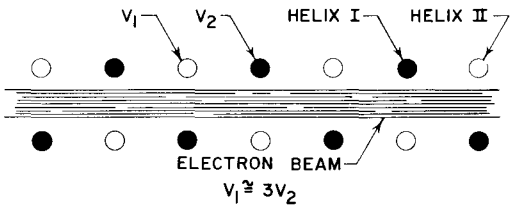


Figure 38. A Bifilar-Helix Electrostatic Focusing Arrangement

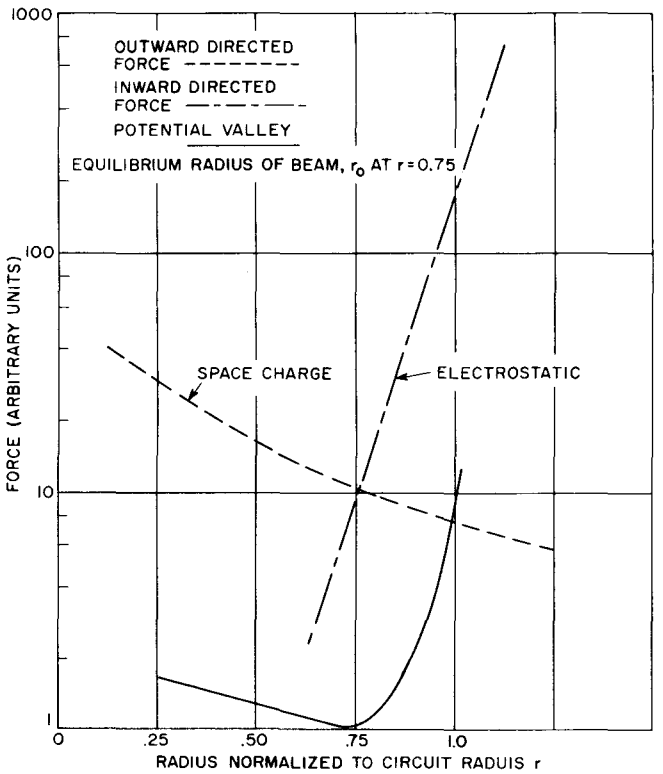


Figure 39. The Relative Amplitudes of the Forces Acting on an Electron at a Given Distance from the Beam Axis in the Bifilar-Helix Electrostatic-Focusing Arrangement

The chart by Blattner and Vaccaro shown in Fig. 40 relates the ratio of beam-to-helix diameters, the helix-voltage ratio, and the maximum perveance which can be focused. The second helix wire in the bifilar helix does not affect the rf characteristics of the helix, and both the phase velocity and impedance are unchanged from that of a unifilar helix having the same pitch angle and diameter.

The original form of electrostatic focusing by means of bifilar helices employed a stiffer form of focusing than that which has just been described. The gun region of the tube was immersed in a magnetic field which was abruptly terminated as the electron beam entered the bifilar helix region. By Bush's theorem, this caused the beam to rotate rapidly about the axis as it left the magnetic field and produced outward forces based on both centrifugal forces and space-charge forces. These forces were then counterbalanced by the inwardly-directed force from the electric field of the bifilar helix. The force diagram shown in Fig. 41 indicates that this arrangement provides a stiffer form of focusing which tends to keep each electron on its original radius. This type of focusing was demonstrated by Chang. [16]

Another scheme proposed and demonstrated by Chang[17] to improve the stiffness of purely electrostatic focusing is the use of an inner bifilar helix concentric with the outer one, but having nonpropagating characteristics; a hollow beam is directed between the two pairs of helices. The force diagram shown in Fig. 42 indicates the resulting forces on the electrons. As seen from this figure, this is a very stiff form of focusing in that the restoring forces are large. Relatively high perveances per square can be focused with good beam definition. Because of the complexity of the device and because of the poor heat dissipation properties of the inner helix, this has only been used on one experimental tube.

Folded-Waveguide Structure. Fig. 43 shows a typical cross section of a folded-waveguide slow-wave circuit in which focusing spades have been inserted between the drift tubes of the slow-wave circuit. By isolating these focusing spades from the body of the tube and applying a potential either lower or higher than the body potential with a ratio between the two voltages of about 3:1, a 2.5 micro-perveance beam can be focused. One such tube has produced 10 kilowatts of peak power at X-band. Details of the design and construction of this tube are given in several papers by Siekanowitz and Vaccaro[18] and Belohoubek[19].

The basic philosophy of this design is similar to that of the parallel-flow gun in that the spades and drift tubes, which are shaped to provide the proper potential along the beam boundary, simulate the flow of electrons between two infinite planes. Fig. 44 compares the voltage shape actually obtained with the ideal.

Radial Electric Field. In Harris flow is an example of focusing with radial electrostatic fields; a cylinder or wire is placed coaxially inside the slow-wave structure and a hollow beam is shot between the cylinders from a gun immersed in a magnetic field. The mag-

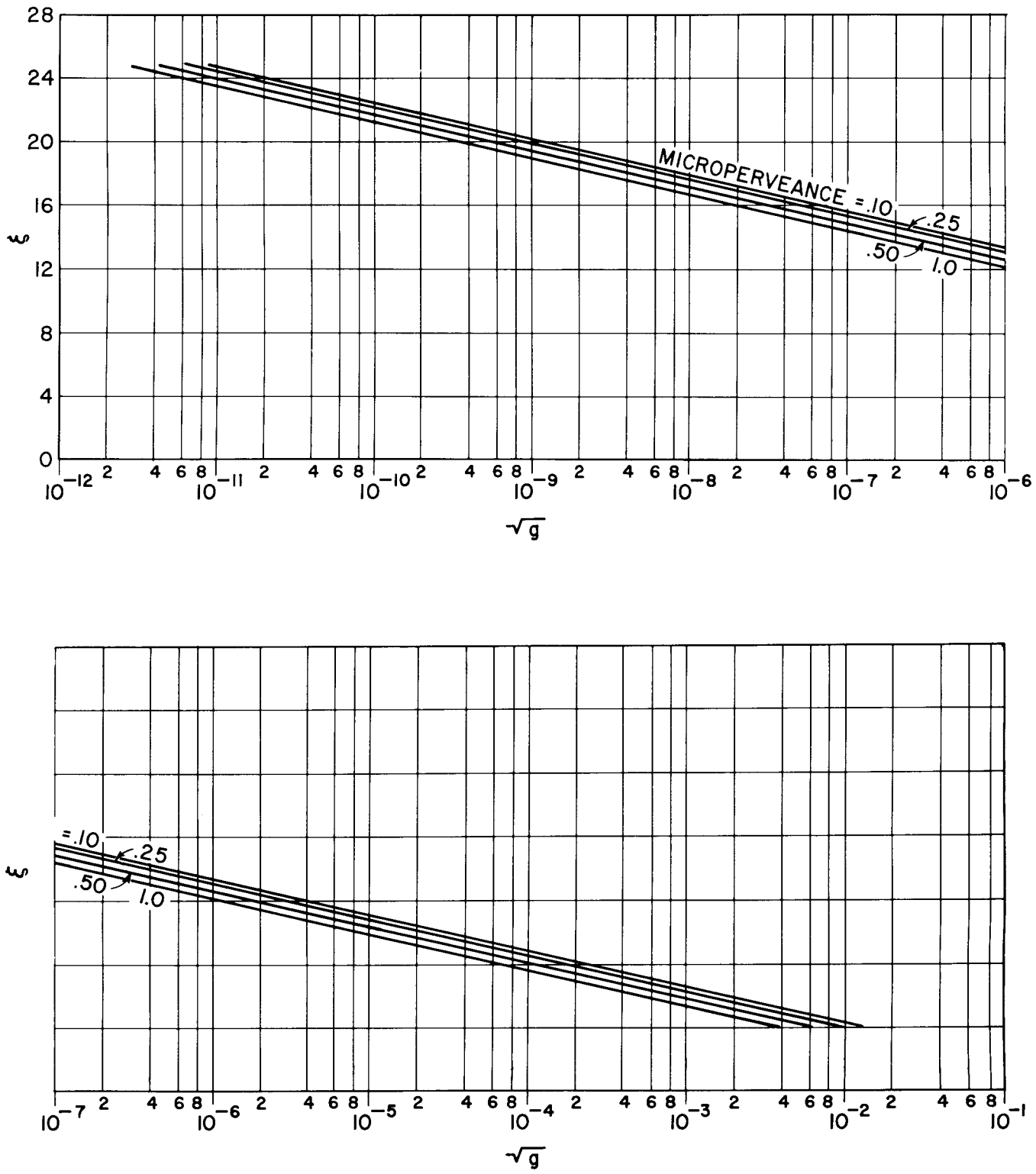


Figure 40. Design Chart for the Bifilar-Helix Electrostatic Focusing Structure

netic field is abruptly terminated as the electrons leave the gun region and enter the radial electrostatic field region. By Bush's theorem, the beam begins to rotate about the axis with a certain angular velocity with the result that there are two forces tending to push the beam outward: the centrifugal force of the rotating

electrons, and the space-charge forces in the beam itself. These outward forces are counteracted by the radial electric field produced by operating the inner cylinder at a higher potential than the outer cylinder, as shown in a cross section of a typical Harris flow traveling-wave tube in Fig. 45; Fig. 46 shows the resulting force diagram. Because there is less radial dependence of the inward electrostatic force than there is for the inward magnetic force in the case of Brillouin flow, Harris flow is not as stiff and stable as Brillouin flow. Only a few experimental tubes have been made using this focusing scheme, because of the dissipation problems on the inner cylinder and the problem of generating a magnetic field in the gun region and cutting it off sharply as the interaction region is reached.

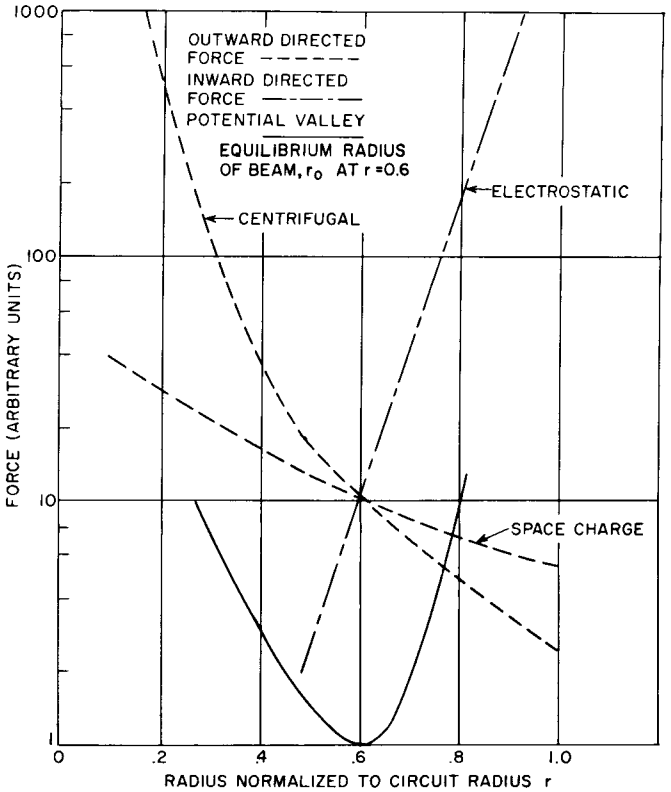


Figure 41. Diagram of the Forces Acting on the Electron Beam in the Bifilar-Helix Electrostatic Focusing Arrangement when the Electron Gun is Immersed in a Magnetic Field that is abruptly Terminated at the Entrance to the Helix Region

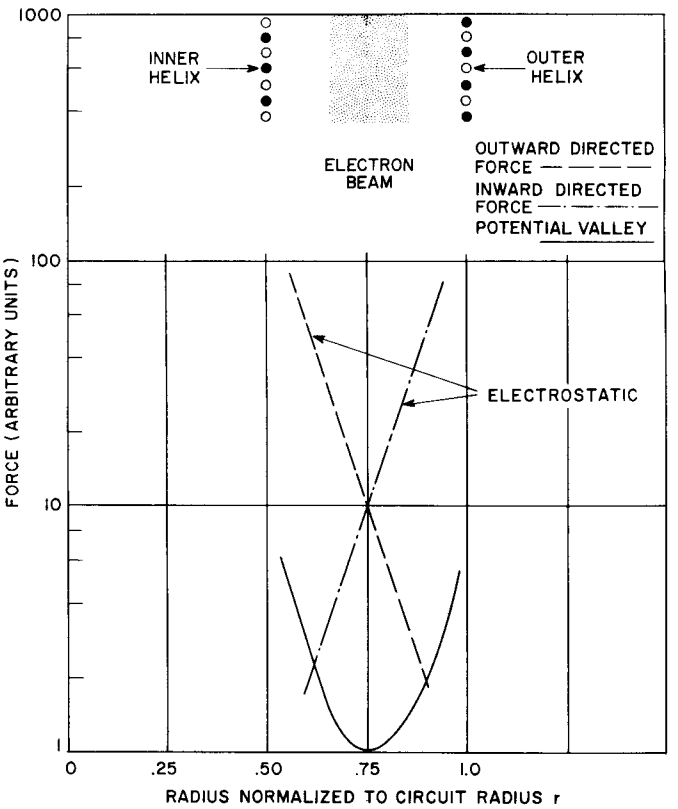


Figure 42. Diagram of the Forces Acting on an Electron in a Hollow Beam Propagating the Annular Space between two Concentric Bifilar Helices

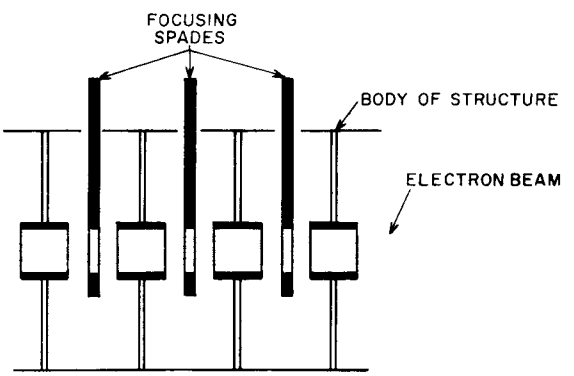


Figure 43. Cross Section of an Electrostatically-Focused Folded-Waveguide Slow-Wave Circuit

COUPLERS

The rf couplers of a traveling-wave tube are probably the nearest thing to a necessary evil on a traveling-wave tube. They frequently take up space urgently needed for the magnetic focusing, limit the bandwidth of the tube, or require extra seals which are likely to develop leaks. The basic function of the coupler is to provide a means for passing rf energy from an external line to the slow-wave structure and then from the slow-wave structure back to the output line. In doing so, the impedances of the input and output lines must be matched to the impedance of the slow-wave structure.

Coupling to Helices

Direct Connector. The simplest and most direct way of coupling the rf energy into and out of the helix is by means of a direct connection to a coaxial line, as shown in Fig. 47. The impedance from the helix to the coaxial line is matched by tapering the entrance of the helix and providing a close shield about the helix. This type of match is very broad band and usually a VSWR of less than two can be maintained over a band of an octave or more. However, it is fairly difficult to construct because a wire must pass through the envelope of the tube and some form of tuning device is needed to

improve the impedance match. A very low VSWR, e.g., below 1.2:1, is difficult to achieve either over a wide frequency range or at one specific frequency. However, in certain types of tubes, specifically backward-wave amplifiers or oscillators, in which coupling is difficult because of the type of propagating mode set up on the helix, this method of coupling is still in wide use.

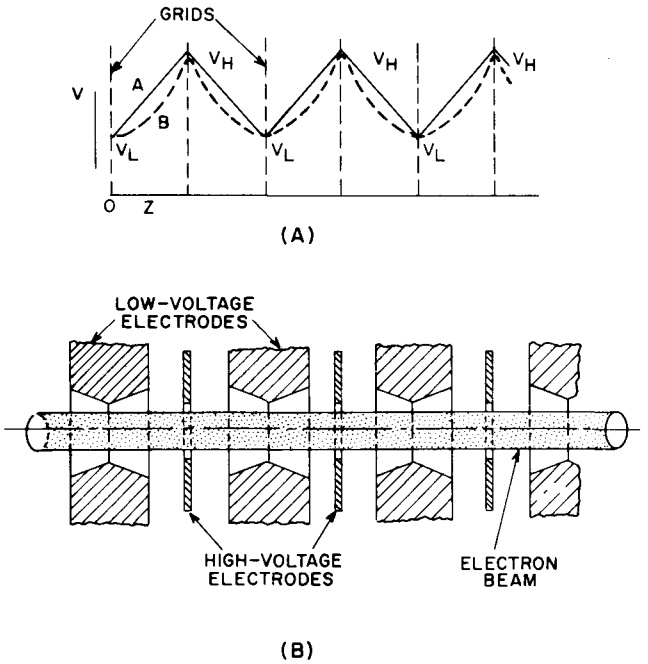


Figure 44. Electrostatically-Focused Folded-Waveguide Slow-Wave Structure: (A) Comparison of the ideal and actual shape of the focusing field; (B) Cross section of the structure

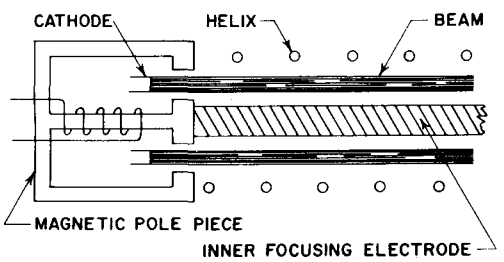


Figure 45. Cross Section of the Electron-Gun and Helix Regions of a Typical Harris-Flow-Focused Traveling-Wave Tube

Cavity Coupler. The cavity coupler, because of its large size and relatively narrow bandwidth, is not consistent with today's small light-weight broadband traveling-wave tubes. However, it still has some limited use in specific applications such as narrow-band low-noise traveling-wave tubes in which the coupler loss must be reduced to a bare minimum. Fig. 48 shows the cavity coupler used on the RCA-6861 low-noise traveling-wave tube. The pertinent dimensions are shown on the drawing; for design purposes, it is recommended that dimensions be scaled to other frequencies from the dimensions given in the drawing.

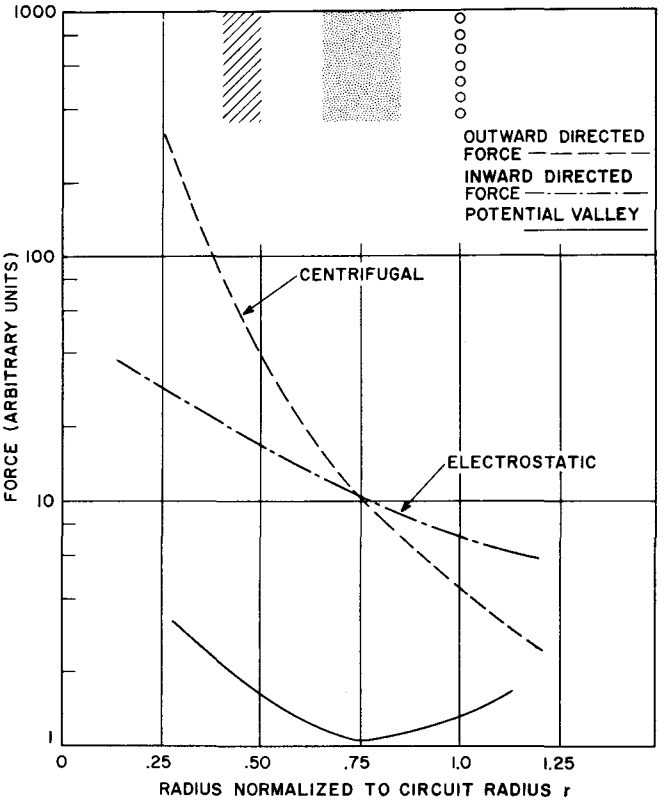


Figure 46. The Focusing Forces Acting on the Electron Beam in a Harris-Flow-Focused Traveling-Wave Tube

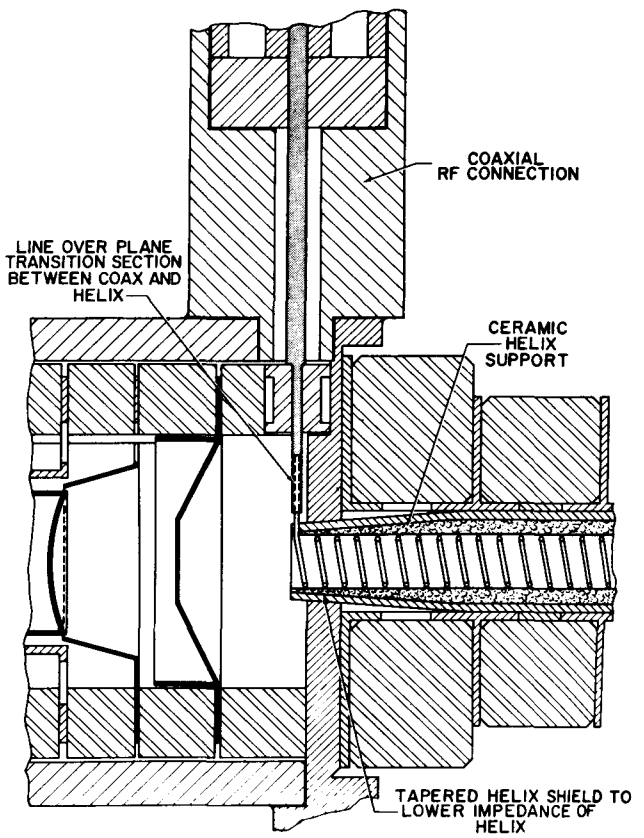

COAXIAL RF CONNECTION
LINE OVER PLANE TRANSITION SECTION BETWEEN COAX AND HELIX
CERAMIC HELIX SUPPORT
TAPERED HELIX SHIELD TO LOWER IMPEDANCE OF HELIX

Figure 47. Cross-Sectional Diagram Showing the Direct-Coupling Method of Coupling RF Energy to the Helix

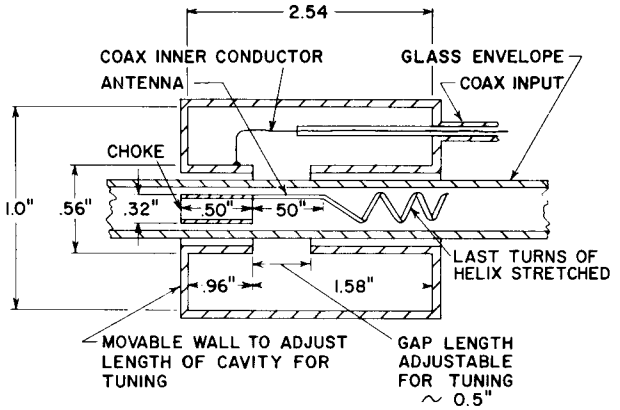


Figure 48. Cross Section of the Cavity Coupler Used in the RCA-6861 Traveling-Wave Tube

In this type of coupling, the helix is terminated by a short antenna followed by a choke section. The last three or four turns of the helix are stretched to provide a smooth transition into the antenna. Surrounding the antenna is an open portion of a fairly broadband and low-Q cavity. The transition of the cavity to a coaxial line is made as shown in Fig. 48.

Several adjustments can be made to achieve the lowest possible coupler loss over a specific frequency range; for example, the width of the gap of the cavity or the over-all length of the cavity can be changed. Over relatively narrow frequency ranges of 400 to 500 megacycles at S-band, a very low VSWR in the order of 1.1:1 can be regularly achieved.

Waveguide Couplers. In the waveguide method of coupling, the helix is terminated in exactly the same manner as in the helical-coupler method with several stretched turns, an antenna, and a choke section. The impedance of the helix is matched by making the waveguide thinner by a factor of 3 or 4, either by a smooth transition or by a multiple-step transition. A hole cut through the waveguide, as shown in Fig. 49, is placed over the antenna. One end of the waveguide is terminated with an adjustable short and the other end goes through the transition to the standard-height waveguide. Bandwidths of more than one-half octave have been obtained with this type of coupling, with a VSWR of less than 2:1. Over narrow-frequency ranges, by adjusting the position of the short and adding additional tuning screws, low values of VSWR of less than 1.1:1 can be maintained. Fig. 50 shows the dimensions of a typical waveguide coupler for an X-band traveling-wave tube; couplers for other frequency bands can be scaled from these dimensions.

Waveguide-type couplers are most commonly used in the frequency ranges of C-band, X-band, and above for two reasons. First, at the lower frequencies, the waveguide is large and bulky (frequently much larger than the tube itself). In such cases, some form of coaxial coupling is preferred. Second, at C-band and X-band, where the helices are very small and supported on ceramic rods, the diameter of the envelope of the tube can be several times the diameter of the helix. Under these conditions, the more commonly used helical couplers, to be described later, cannot be used, and waveguide couplers are the preferred alternative. At any frequency, where very high cw or pulsed power is generated, the low-loss waveguide is preferred to coaxial types.

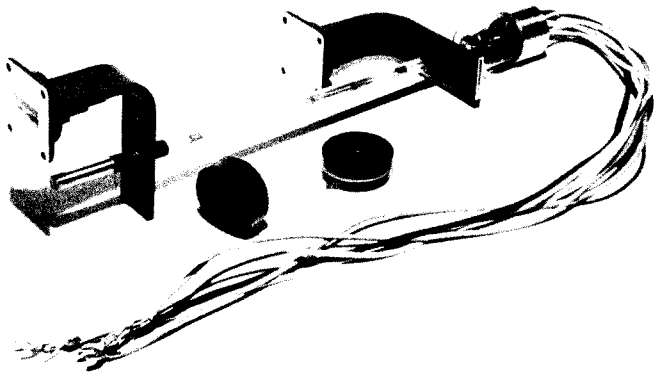


Figure 49. Arrangement Used in the Waveguide Method of Coupling Energy into and out of the Helix

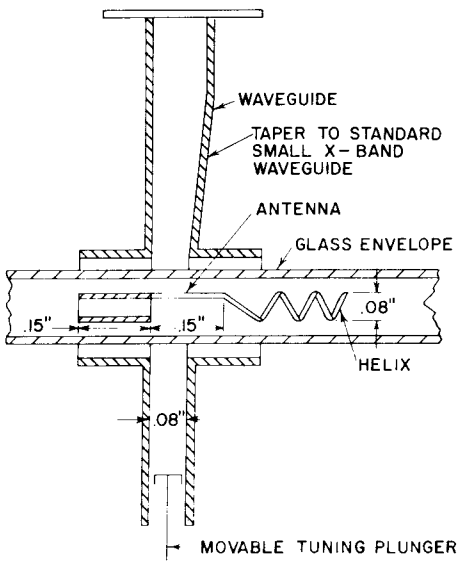


Figure 50. Cross Section of a Typical Waveguide Coupler for an X-Band Traveling-Wave Tube

Helical Couplers. In recent years, most of the tubes designed and in production employ helical couplers. The helical coupler is extremely broadband; excellent coupling can be obtained over more than an octave of bandwidth. They take up little space and can be easily built inside a periodic permanent magnet stack. Unlike the direct connection, which has a roughly equivalent bandwidth, the coupled-helix coupler does not require a break in the vacuum envelope.

The bandwidth of the coupled-helix coupler is much wider than that obtained with a cavity coupler or with a waveguide coupler and is comparable to that obtained with a direct connection. The helical coupler can be made with a lower VSWR than the direct connection, typically less than 1.5:1 over an octave. But the helical

coupler will add several inches to the length of the tube.

The helical coupler occupies less volume than other types and thus causes little interference with the location and operation of the focusing magnets. In addition, the helical coupler can couple to the bifilar helix used in electrostatic focusing without coming in direct contact with either of the helices. This can be very important in electrostatic focusing where the helices can have a potential difference of several thousand volts.

The theory and design of coupled-helix couplers has been thoroughly treated by Cook, Kompfner, and Quate.[20] Only the basic theory and some construction details are presented here. The theory of operation is simple, as shown in Fig. 51. Two helices having different diameters are wound concentrically with one wound in the opposite direction to the other. If an rf signal is introduced on the helix at the left-hand side, the signal propagates down the helix in the right-hand direction. Because the field of the helix extends into the region occupied by the other helix, some of the energy is transferred to the other helix. If the phase velocity of the two helices are the same and have the proper diameter ratio, a beating phenomenon occurs which is similar to the mechanical analog of the coupled pendulum. After traveling a short distance, the energy, which is initially all in the outer helix, is transferred entirely to the inner helix. At a later point, it is transferred back entirely to the outer helix. If, however, the outer helix is terminated at the point at which all the energy is on the inner helix, the energy is, in effect, transferred or coupled from a coupling helix to the tube helix. If the process is reversed, an output coupler for a traveling-wave tube is formed.

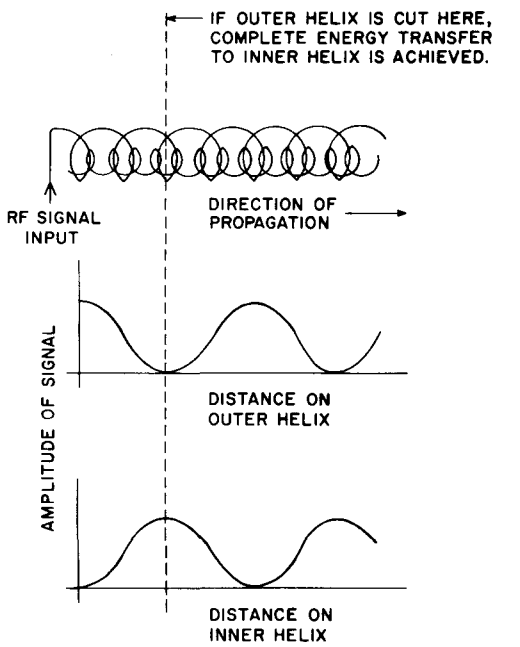

IF OUTER HELIX IS CUT HERE, COMPLETE ENERGY TRANSFER TO INNER HELIX IS ACHIEVED.
RF SIGNAL INPUT
DIRECTION OF PROPAGATION
AMPLITUDE OF SIGNAL
DISTANCE ON OUTER HELIX
DISTANCE ON INNER HELIX

Figure 51. Diagram of the Interaction in the Coupled-Helix Coupler

Typical construction of such a coupler is shown in Fig. 52. A 50-ohm coaxial line having a center conductor wound into the form of a helix which is surrounded by a close-spaced metal shield. With the proper spacing between this metal shield and the coupling helix, the impedance of the line-over-plane formed by the coupling helix in the shield can be made to match that of the coaxial line, and no impedance mismatch occurs at the transition. The coupling helix then couples with the tube helix as described previously. From the referenced article, it is possible to find the optimum diameter ratio for the two helices for a given frequency range and calculate the length of the coupled helix. However, in practice, the length of the coupling helix is normally one to two times the calculated value. The easiest means to determine the actual length is to make an overlength coupler and then gradually cut off turns until the minimum VSWR is achieved.

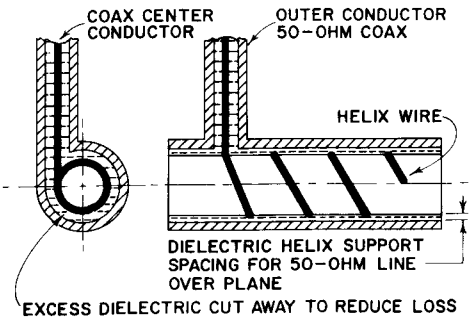

COAX CENTER CONDUCTOR
OUTER CONDUCTOR 50-OHM COAX
HELIX WIRE
DIELECTRIC HELIX SUPPORT SPACING FOR 50-OHM LINE OVER PLANE
EXCESS DIELECTRIC CUT AWAY TO REDUCE LOSS

Figure 52. Cross Section of a Coupled-Helix Coupler

The optimum spacing ratio between the inner and outer helices for broadband operation is about 1.6:1. At higher or lower ratios, the bandwidth of the coupler becomes more restricted. At higher ratios, e.g., 2:1, the length of the coupler required for complete transfer of energy may become excessive. For this reason, in C-band and X-band tubes in which the relatively small fragile helix is supported on three ceramic rods and the outside of the tube envelope may be several times the helix diameter, helical couplers are not generally used. However, at low C-band, S-band, and L-band using larger and more physically rugged helices, which can be held in precision glass or imbedded in fluted glass, the outer diameter of the envelope may be only 1.3 times the helix diameter and helical couplers may be used.

Coupling to Other Types of Slow-Wave Structures

Means of coupling to other types of slow-wave structures are nearly as numerous as the types of slow-wave structures. In the case of a coupled-cavity or folded-waveguide structure, the last cavity opens up into a waveguide, usually of reduced height, which is then tapered to the size of a standard waveguide, as shown in the cross section in Fig. 53. In the general design procedure, a waveguide is tapered down to the size of the final cavity and the impedance presented to the waveguide is measured at various frequencies across the passband of the slow-wave structure. This data is then plotted on a Smith Chart and the required amount of added inductance or capacitance is determined. Normally, the addition of an inductance, as shown in Fig. 53, is required as an inductive post where the waveguide joins the first cell of the structure.

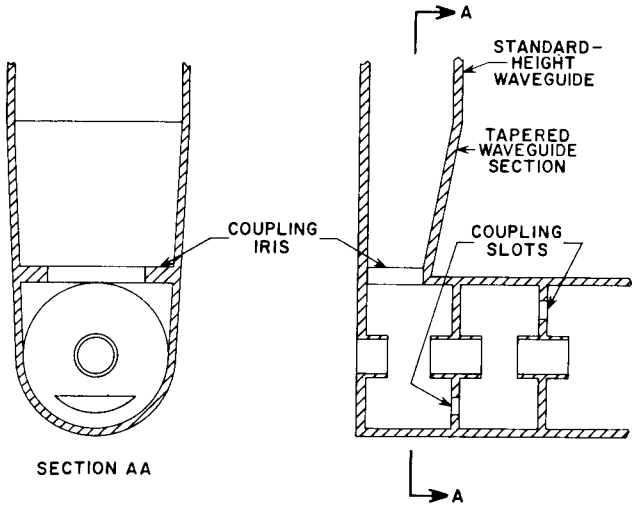


Figure 53. Method Used for Coupling Energy into a Folded-Waveguide Slow-Wave Structure

ATTENUATORS

The first traveling-wave tube, built by R. Kompfner[21] in England, did not have an attenuator and, consequently, was an extremely unstable device. Reflections from the output or from the load could be reflected back through the tube, be re-reflected from the input or the generator, reamplified in the tube and, depending upon whether in or out of phase with the original signal, could either cause oscillations at some frequencies or cancellation of output at other frequencies. It remained for Pierce and Field[22] in this country to add an attenuator in the center of the tube to isolate the input section from the output section and create a stable high-gain wide-band amplifier. For stability, the value of the attenuation should be about 10 db greater than the gain of the tube.

In the same way that lack of an attenuator can cause oscillations, reflections from the attenuator itself can produce the same effect. For instance, in the output section of a traveling-wave tube, the gain may be as great as 30 db. If a mismatch exists in the output coupler or in the load with a VSWR of 2:1, the power reflected back down the helix from the mismatch is 10 db less than the output signal. If the match at the input of the attenuator has a 2:1 VSWR, the power being re-reflected from the attenuator is 20 db less than the output signal. However, because the gain of the output section is 30 db, the signal reflected from the attenuator is amplified and emerges from the tube at a greater amplitude than the original signal. This condition leads to unwanted oscillation.

Many parameters control the placement, the amount, and the type of attenuation used. Among these are the amount of power handled by the attenuator, the efficiency required from the tube, how the output power is to vary with the input power near the saturation region, and the type of slow-wave structure used. These parameters are discussed below in connection with the various types of attenuators.

Attenuators for Helices

In helix-type traveling-wave tubes, the design of the attenuator frequently depends on the type of helix support used. For tubes having rod-supported helices, the attenuation must be applied inside the vacuum envelope, usually on the rods. Helices supported directly inside a glass envelope may use an attenuator external to the vacuum envelope in the form of a coupled-helix attenuator or lossy material painted on the outside of the glass envelope.

Attenuators on Rod-Supported Helices. In tubes having rod-supported helices, a carbon suspension, Aquadag, can be sprayed on the rods at the proper location to supply the desired attenuation. (Usually, at the ends of the attenuator, the amount of carbon sprayed on is tapered in density or geometrically shaped to provide a good match looking into the attenuator.) An excellent discussion of the techniques used in applying attenuation to rod-supported helices is given in an article by Laico, McDowell, and Master.[23]

This type of attenuation is usually limited to low average-power traveling-wave tubes because reflected power absorbed by the attenuator is dissipated inside the vacuum envelope and can cause overheating or outgassing of the rods or the helix. One drawback to this form of attenuator is the lack of adjustment to either the placement or the amount of the attenuation after the tube has been constructed and processed. Particular difficulty results in cases in which the processing of the attenuation is not closely controlled, and drastic changes in the amount of attenuation occur during tube processing.

The primary advantage to this type of attenuation is its effectiveness over a broad frequency range, usually extending for an octave above and below the operating range of the tube. This wide range can be particularly important in suppressing certain out-of-band oscillations, especially those of the backward-wave type which may occur at nearly twice the highest operating frequency of the tube. This type of compensation is frequently not possible with other types of attenuators.

Coupled-Helix Attenuators. One means of overcoming the power handling problems inherent in an internal attenuator is to use a coupling helix (described previously) to extract the energy from the tube helix to a point outside the tube envelope. Two types of coupled helix attenuators are commonly used — a lossy-wire coupling helix or a lossless-wire coupling helix imbedded in a lossy dielectric.

The lossy-wire coupling helix is made by winding a coupling helix of Karma or Baker alloy wire having a diameter of 0.0005 inch. The helix is held in shape by winding the wire on a cellophane-tape mandrel and then covering it with another layer of cellophane tape. If the tube must operate at high temperatures, Teflon tape is substituted for the cellophane tape. This type of attenuator is quite effective and 10 to 20 db of loss per wavelength can be obtained. The number of turns per inch and the diameter of the coupling helix are determined as described under COUPLERS.

The second method, having much greater power-handling capability, uses a relatively lossless coupling

helix wound with 0.005 to 0.010 inch tungsten wire for rigidity. This coupling helix is then placed on a cylinder of material having a loss of about 2000 ohms/centimeter. The cylinder with the coupling helix is then placed over the precision glass supporting the tube helix.

The lossy cylinder is usually a porous ceramic (Alsimag 222) impregnated with carbon by either the methane cracking process or by soaking the ceramic in a 10:1 water-sugar solution and firing in hydrogen to reduce the sugar to carbon. Because it is of metal-ceramic construction and located outside the tube envelope, this type of attenuator can handle many watts of reflected power without damaging the tube.

Both these types of coupled helix attenuators have high reflection coefficients. A mismatch of 1.5:1 or 2:1 VSWR at the input of this type of attenuator is not uncommon. Some type of buffer attenuation must be provided at both ends of the attenuator to reduce reflections. This is generally provided by painting an Aquadag arrow on the outside of the helix bulb or by pasting on arrows cut from carbon-impregnated paper. A typical coupled-helix attenuator with Aquadag arrows is shown in Fig. 54.

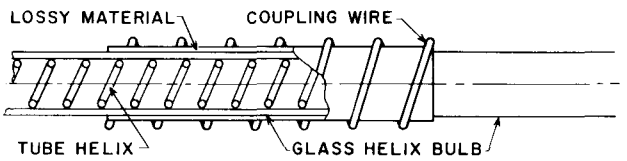


Figure 54. Cross Section of a Typical Coupled-Helix Attenuator

Coupled-helix attenuators are especially useful on bifilar-helix electrostatically-focused traveling-wave tubes. Aquadag-coated rod attenuation would short out the helices, but the coupled helix attenuator, which makes no physical contact cannot cause shorting.

Severed Helix. In general, the attenuator must be at least 2 or 3 wavelengths long to provide the usually required attenuation of 50 to 60 db. At normal frequencies (S-band and above), the length involved is relatively short, especially in low-voltage tubes. However, for high-voltage tubes at S-band and for tubes in the lower frequency ranges, three or four helix wavelengths can require several inches of length and in some high-power low-frequency tubes, the attenuator may add as much as a foot to the over-all length of the tube. In these cases, it is frequently desirable to sever the helix in the center, usually by inserting a drift tube between the input and output section of helix, and terminate the helices at the drift tube with just sufficient attenuation to prevent reflections from the short caused by the drift tube. In this way, 60 to 90 db of attenuation can be provided over shorter distances.

Attenuators for Other Types of Slow-Wave Structures

The problems involved in the design of attenuators for high-power tubes are even more complex than those involved in the helix-type tube. For optimum efficiency, the attenuator should be as short as possible. In the case of the folded-waveguide or other types of coupled-cavity structure, it is desirable that not more than one or two cells be used for the attenuator. The circuit is severed at the point where the attenuation is desired by sealing off the coupling slots between the cavities. This effectively prevents the propagation of the wave in either direction past this point and immediately introduces an isolation in the order of 50 to 60 db.

A lossy substance introduced into the cavities on each side of the break provides a good match and absorbs the incident power. The material used is a porous ceramic described above. Generally, the lossy material almost entirely fills the cavity, although the exact shape must be determined experimentally to find the configuration which provides the lowest VSWR.

When the power is so great that it cannot be handled by an internal attenuator, waveguide couplers are used on either side of the break to take the reflected rf power out of the tube when it may be dissipated in an external load. To reduce the VSWR presented to these attenuators, several cavities (before and after the attenuator) are lightly sprayed with Aquadag to provide some buffer attenuation.

Attenuators for other types of slow-wave structures follow essentially the same general design as for the helix or for the folded waveguide. Some lossy material, either Aquadag or ceramic impregnated with carbon, is placed on or near the slow-wave structure and some form of taper is provided to maintain a good match at the input of the attenuator.

COLLECTOR

In many low-power traveling-wave tubes, almost any piece of copper is satisfactory as a collector if operated about 100 volts above the helix potential to prevent secondary electrons from returning to the helix from the collector. However, for tubes above the one-watt level, the collector design must be more carefully considered. The power which the collector must dissipate may be considerable because of the low efficiency of most traveling-wave tubes. In a 100-watt tube, the beam power can range from 500 to 1000 watts.

Depressed-Collector Operation

Because of the low-helix interception, nearly all the dc power in the tube is dissipated at the collector. Thus, any reduction in the collector potential at which the beam is collected can materially increase the operating efficiency of the tube.

One of the primary disadvantages of traveling-wave tubes, particularly in the high-power ranges, has been their relatively low efficiency compared to crossed-field or "M"-type devices. Thus, traveling-wave tubes have been used only when equipment designers were forced to use them because of their tremendous bandwidth. As the techniques of collector depression become more advanced and more widely used, the traveling-wave tube, operating at efficiencies of 40 to 50 per cent, will be suitable for many equipments where amplitrons and other crossed-field devices are presently used.

The efficiency of an intermediate power tube such as a 100-watt traveling-wave tube, which would normally operate at about 10 per cent efficiency, can be doubled or tripled simply by reducing the collector voltage and collecting the spent electron beam at voltages far below the helix voltage. However, several factors have to be considered before this type of operation is readily possible. For instance, when the collector is operated at a much lower potential than the helix, secondary electrons generated by the beam striking the collector, are drained from the collector to the helix or slow-wave structure, and cause local overheating. Another limit to the lowest value at which the collector can be operated is set by the space-charge depression in the center of the beam. The effects of space-charge depression and its limitation on the amount by which the collector can be depressed have been discussed in an article by Wolkstein.[24] The limit is dependent on the amount of beam current and on the ratio of the beam diameter to the inner diameter of the collector. For example, a 0.25-ampere beam in a collector with a diameter twice the beam diameter, can be depressed no lower than 550 volts. Fig. 55 is a design chart from which the space charge depression of any beam can be determined.

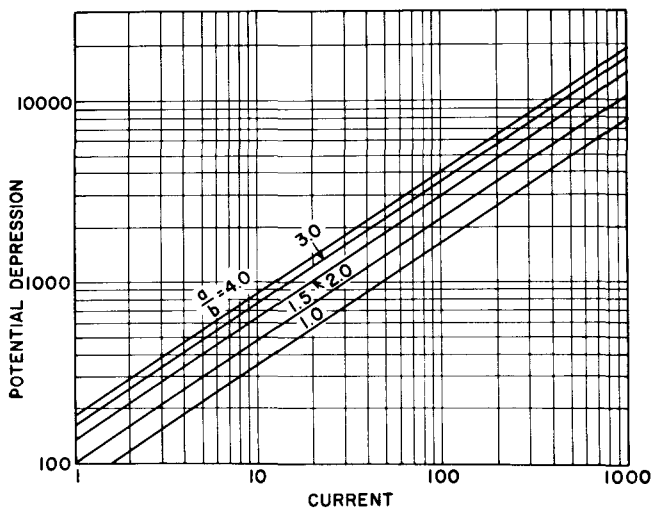


Figure 55. Design Chart Used to Determine the Space-Charge Depression of an Electron Beam

Two-Stage Depressed Collector

An additional factor which determines the minimum potential to which the collector may be depressed is the velocity spread of the electrons after they have given up their energy to the rf field. Because the rf energy comes from the kinetic energy of the electrons, the beam is slowed down and a higher potential must be used to collect the electrons. Not all electrons are slowed down by the same amount and, as a result, there is a considerable velocity spread in the electron beam. An article by F. Sterzer[25] describes this effect in detail. His solution to the problem of optimum collector depression is to divide the collector into two segments as shown in Fig. 56. In this way, the few electrons which have given up the most energy to the rf wave and, therefore, require a higher collector voltage are collected on the first segment. The remainder of electrons are then collected on the second segment which collects the bulk of the beam at a much lower voltage. With this approach, he was able to obtain an efficiency of over 45 per cent in a 100-watt S-band traveling-wave tube in which the efficiency would otherwise be only 15 per cent.

Secondary-Emission Suppression

The secondary electron emission from the collector to the helix may be reduced by reducing the emission coefficient of the collector. A second and more efficient means is to trap the secondary electrons in the collector by making a small entrance to a large collector — essentially a sort of Faraday Cage. The magnetic field must be terminated at the entrance of the collector to prevent the secondary electrons from following the lines of flux back into the helix region. Fig. 57 shows a collector of this type which has been used on several traveling-wave tubes with a very small number of returned electrons, even with the collector depressed to near the theoretical limit.

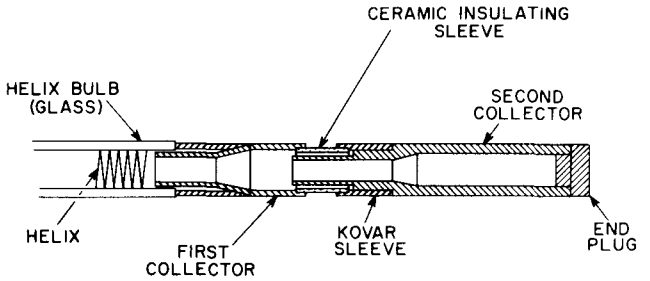


Figure 56. Cross Section of a Two-Stage Collector Assembly

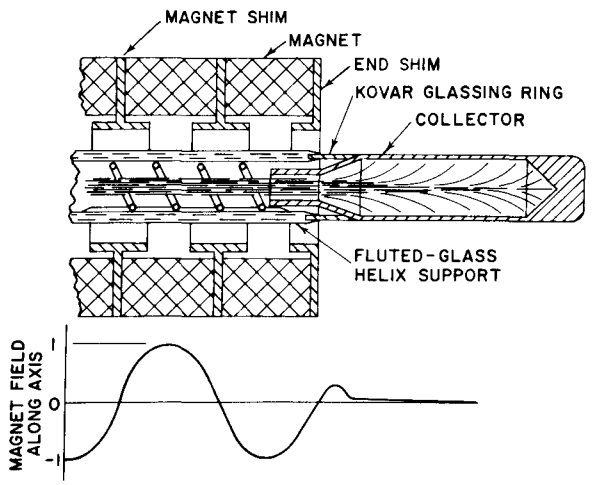


Figure 57. Cross Section of a Magnetically-Shielded Collector Assembly

Collector Cooling

Cooling of the collector is handled by relatively conventional means — either by heat transfer to nearby massive objects, by forced air cooling, or by some form of liquid cooling.

APPENDIX

DESIGN PROCEDURE FOR TRAVELING-WAVE TUBES

As an illustrative example of the procedures followed in the design of a traveling-wave tube, let us assume that a customer has requested an amplifier having the following characteristics:

Frequency range	2.0-4.0 Gc
Power output	1.0 kilowatts
Duty cycle	0.01
Gain (at 1.0 kw)	33 db

The broad bandwidth requires a helix type circuit. With the low duty cycle resulting in an average power of 10 watts, the helix is sufficiently robust to handle the power. Focusing will be supplied by periodic permanent magnets. Because the customer did not specify the voltage, we will calculate tube designs for a range of voltages and then interpolate between these to determine the optimum design based on available electron guns, magnetic field, and size of the tube.

The items given above, which include the midband frequency of 3.0 Gc, are normally specified by the customer and thus are not under the control of the designer.

The following three parameters do not change appreciably from tube to tube and for a first approximation may be chosen as follows:

Efficiency	15 per cent
Dielectric Loading Factor	0.85
γ_a	1.5

The efficiency is roughly equal to 2C (where C is Pierce's gain parameter and the validity of the original assumption can be checked later in the calculation when C is determined. The dielectric loading factor can be estimated depending upon the type of helix support to be used but cannot be checked until a helix assembly has been constructed and tested. The midband value of γ_a was chosen because of the wide bandwidth required.

The following two items are entirely at the discretion of the designer:

Helix phase velocity, v_p	3.0 kilovolts
	5.0
	8.0
Ratio of beam to helix diameter	0.5

The effect of the beam-helix diameter was discussed in the text while the effect of the various helix voltages will be noted as a result of the calculations.

In the following discussion, the terms voltage and velocity will be used interchangeably. They are related by the equation

$$\frac{v}{c} = \frac{\sqrt{V}}{506}$$

where v is the velocity, V is the voltage, and c is the speed of light.

GENERAL TUBE DESIGN

Parameter	Units	Design 1	Design 2	Design 3
Midband frequency	Gc	3.0	3.0	3.0
Wavelength	centimeters	10.0	10.0	10.0
Helix phase voltage, V_p	volts	3000	5000	8000
Helix phase velocity v_p/c		0.1085	0.140	0.1755
Beam current $I_o = \frac{P_{out}}{\eta V_p}$	amperes	2.22	1.33	0.835

The space charge potential depression is the difference between the voltage applied to the helix and the actual velocity of the electrons. This difference is caused by the negative charge of the electrons as explained in the text.

Parameter	Units	Design 1	Design 2	Design 3
Space charge potential depression (See Fig. 55)	volts	2300	1650	1200

In order for the electrons to give up energy to the rf wave, the electrons must travel faster than the cold phase velocity of the helix. This difference is approximately equal to 3/4 of the space charge potential depression.

Parameter	Units	Design 1	Design 2	Design 3
Difference between v_p and u_o	volts	1700	1250	900
Voltage corresponding to electron velocity, V_o	volts	4700	6250	8900
Power supply voltage, V_{ps} (electron velocity plus space charge depression voltage)	volts	7000	7900	10100

The proper voltage (i.e., V_p, V_o, or V_{ps} and the corresponding phase velocities) must be used at the appropriate occasions below as noted.

Parameter	Units	Design 1	Design 2	Design 3
$\gamma_a = \frac{2\pi a}{\lambda}\frac{506}{\sqrt{V_p}}$ [See Eq. (17)]		1.5	1.5	1.5
$ka = \frac{2\pi a}{\lambda}$		0.163	0.21	0.263
Helix radius, a (Calculated from equation for γ_a)	inches	0.102	0.132	0.165
$F(\gamma_a)$ (See Fig. 15)		2.62	2.62	2.62
IRF (See Fig. 16)		1.075	1.075	1.075

Parameter	Units	Design 1	Design 2	Design 3
F [See Fig. 17)]		0.45	0.42	0.38
K - use V_p [See Eq. (25)]	ohms	43.2	31.3	22.5
C - use V_o [See Eq. (7)]		0.172	0.118	0.081

At this point we can calculate the efficiency and check the beam current which was calculated earlier using the approximate efficiency. The efficiency will be twice the value of C but reaches a limit about 0.2. The voltage corresponding to the electron velocity, V_o, should be used, and the power output should be about 1.5 times the minimum required to allow for the drop in power near the ends of the frequency band.

Parameter	Units	Design 1	Design 2	Design 3
η		0.20	0.20	0.16
I_o	amperes	1.6	1.2	1.0

Using these new values of current, the electron velocity and power supply voltage can be recalculated and the value of C recalculated. The helix impedance K will not change because the phase velocity does not change.

Parameter	Units	Design 1	Design 2	Design 3
Space charge potential depression	volts	1850	1500	1350
Difference between v_p and u_o	volts	1400	1100	1000
V_o	volts	4400	6100	9000
V_{ps}	volts	6250	7500	10350
C		0.158	0.115	0.085

We can now recalculate the beam current using the same assumptions as before.

Parameter	Units	Design 1	Design 2	Design 3
η		0.2	0.2	0.17
I_o	amperes	1.70	1.23	0.98

These values of beam current are sufficiently close to the previous set that no further recalculation is required.

Parameter	Units	Design 1	Design 2	Design 3
Q (See Fig. 5A)		1.6	1.6	1.6
QC		0.272	0.197	0.157
B (See Fig. 7)		36	38	39
A (See Fig. 6)		-6.75	-7.0	-7.2
L_{att}		6.0	6.0	6.0

The loss introduced by the attenuator can be read from Fig. 8 assuming the attenuator is a sever in the circuit and the length is known. However, because the attenuator is composed of a finite amount of loss which is tapered at each end, the first assumption is not entirely valid and the exact length is difficult to determine. Thus, an estimate of 6 db attenuation is a reasonable average.

The next step is the determination of the active length of helix required to obtain 33 db gain at a power output of 1.0 kilowatt. It should be noted that the traveling-wave tube equations apply to small signal conditions only in which the rf power output is about 10 per cent or less of the saturated power output. The gain at saturation is normally about 6 db less than the gain under small signal conditions. Thus, in determining the length of the circuit as a function of gain, the small signal gain, 39 db, should be used.

Parameter	Units	Design 1	Design 2	Design 3
N, active length of circuit in helix wavelengths - See Eq. (16)		9.1	11.9	15.7
1, active length of circuit - use V_o - See Eq. (18)	inches	4.7	7.25	11.6
Coupler length	inches	1.2	1.8	2.9
Attenuator length	inches	1.4	2.2	3.9
Gun and collector length	inches	6.0	6.0	6.0
Total tube length		13.5	17.25	24.4

The attenuator length must be about one-third the active length of the circuit to obtain the required attenuation.

The combined length of the input and output helical couplers will normally be about one-fourth of the active length required to obtain 30 to 40 db of gain.

Parameter	Units	Design 1	Design 2	Design 3
r_o, beam radius	inches	0.051	0.066	0.082
B_m, 1.4 B_o - use V_o [See Eq. (28)]	gauss	1450	880	572
L, magnet period for C = 8 [See Eq. (29)]	inches	0.367	0.710	1.33
Gun perveance - Use V_{ps}	μpervs	3.4	1.9	0.81

The power supply helix voltage is used for the gun perveance because we would like the anode to operate at the helix potential.

Some of the more important parameters which have been calculated for the three designs are plotted in Fig. A1 as a function of helix diameter. The only parameter which is limiting is the gun perveance. At

the present time, we do not have a gun with good performance above a perveance of 2.0 and, thus, the smallest helix diameter which could be used is 0.260 inch. A smaller diameter would be attractive because of the shorter length tube which would then be possible as well as the lower operating voltage which is always desirable to the equipment manufacturer. For this tube, however, the present state of the art of the electron gun sets the limit of the design. Tubes at other power levels and other frequencies may be limited by other factors than the gun such as the magnetic field which can be generated by present-day magnet materials.

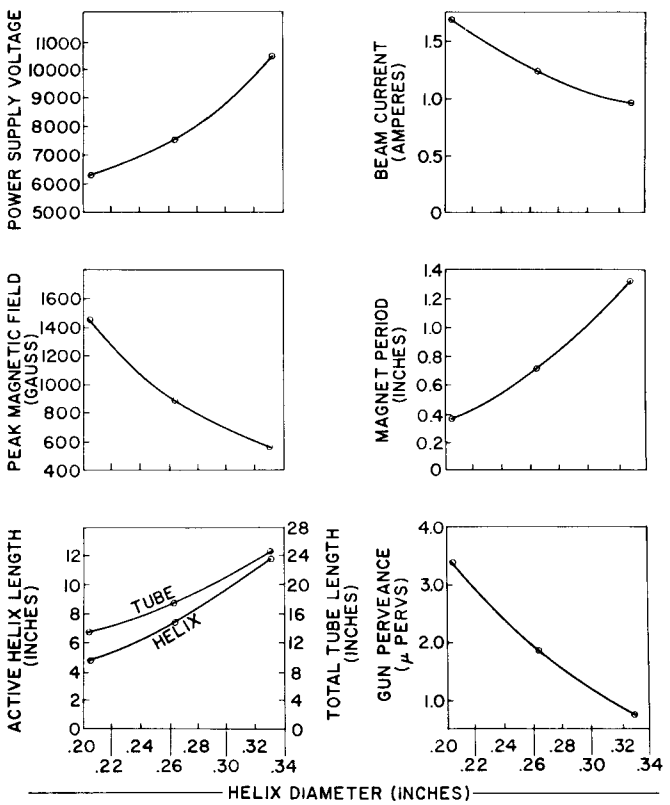


Figure A1. Design Parameters as Functions of Helix Diamater

At this point, the tube design is essentially complete. The major parameters are tabulated below:

Power supply voltage	7400
Beam current	1.25 amperes
Helix diameter	0.260 inch
Peak magnetic field	900 gauss
Magnet period	0.70 inch
Active length of helix	7.0 inches
Total length of tube	17 inches
Electron gun perveance	1.9 μpervs
Phase velocity of helix	5000 volts
Electron velocity	6100 volts
C	0.115

In order to verify that the performance will be uniform across the band, the calculations made above should now be repeated varying the frequency instead of the helix diameter or voltage. The frequencies chosen might be 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, and 4.5 Gc. An effort should be made to estimate the correct phase velocity at each frequency, as shown in Fig. 10, and modified by the dielectric loading factor. To calculate the gain, start with the helix diameter, the phase velocity, and the current at each frequency and compute the electron velocity, the helix impedance C, and then the gain. The results of this calculation will show if the center frequency of the tube corresponds with the required center frequency. If not, γ_a, or the beam-helix diameter ratio, is adjusted and the procedure repeated.

SYMBOLS

- A - Initial loss
- a - Mean helix radius
- B - Growing wave parameter (Fig. 7)
- B_m - Peak magnetic field in a periodic permanent magnet assembly
- B_o - Brillouin magnetic field
- b - Beam radius; also r_o
  - Shield radius
- C - Pierce's gain parameter $C^3 = KI_o/4\,V_o$
  - Constant determining allowable period of periodic permanent magnet stack
- c - Speed of light
- d - Helix wire thickness
- DLF - Dielectric loading factor
- F - Impedance reduction due to DLF (Fig. 17)
- f - Frequency
- I, I_o - Beam current
- i - Square root of -1
- IRF - Impedance reduction factor, actually an increase of helix impedance due to finite size of electron beam (Fig. 16)
- K - Helix impedance
- k - Free space wave constant $k = \omega/c$
- L - Period of loaded waveguide slow-wave structure
  - Period of periodic permanent magnet assembly
  - Loss on helix or other type circuit
- l - Length of circuit
- L_{att} - Loss due to lumped attenuator (Fig. 8)
- N - Length of circuit in guide wavelength $N = 1/\lambda_g$
- P - Helix pitch
  - Perveance, $p = I_o/V^{3/2}$
- Q - Pierce's passive mode parameter
- r - Radial distance from axis of beam, helix, or other slow-wave circuit
- r_o - Beam radius; also b
- u_o - Velocity of electron beam
- V - Voltage
- V_o - Voltage corresponding to velocity of electron beam
- V_{ps} - Voltage applied to helix or other slow-wave structures from power supply
- v - Velocity of growing wave
- v_p - Phase velocity of circuit in absence of electron beam
- α - Attenuation constant
- β - Growing wave phase constant, ω/v
- β_e - Electron beam phase constant ω/u_o
- β_o - Free space phase constant, ω/c
- Γ - Propagation constant of growing wave

Γ_1 - Propagation constant of slow-wave structure

η - Efficiency

γ - $\frac{2\pi a}{\lambda_o} \frac{c}{v}$, radial phase constant

λ_g - Circuit wavelength, $\lambda_o \frac{u_o}{c}$

λ_o - Free space wavelength

ψ - Helix pitch angle, $\tan \psi = \frac{P}{2\pi a}$

ω - Angular frequency, $\omega = 2\pi f$

REFERENCES

1. Pierce, J. R., Traveling Wave Tubes, D. Van Nostrand Co., Inc., New York, 1950
2. Poschl, Van Kalus, "Wellenfortpflanzung langs einer Wendel mit zylindrischem Aubenleiter," A. E. U., Vol. 7, pp. 518-522, 1953
3. Nalos, E.J., "Present State of Art in High-Power Traveling-Wave Tubes," Microwave Jour., Vol. 2, No. 12, p. 31, Dec. 1959; Vol. 3, No. 1, p. 46, Jan. 1960
4. Belohoubek, E., "Propagation Characteristics of Slow-Wave Structures Derived from Coupled Resonators," RCA Review, Vol. 19, No. 2, pp. 283-310, June 1958
5. Nalos, E.J., "Measurement of Circuit Impedance of Periodically Loaded Structures by Frequency Perturbation," Proc. IRE, Vol. 42, No. 10, pp. 1508-1511, October 1954
6. Danielson, W. E., J. L. Rosenfeld, and J. A. Saloom, "A Detailed Analysis of Beam Formation with Electron Guns of the Pierce Type," Bell Syst. Tech. Jour., Vol. 35, No. 2, pp. 375-420, March 1956
7. Pierce, J. R., Theory and Design of Electron Beams, D. Van Nostrand Co., Inc., New York, 1949
8. Helm, R., K. Spangenberg, and L. M. Field, "Cathode Design Procedure for Electron Beam Tubes," Electrical Communications, Vol. 24, p. 105, March 1947
9. Samuel, A. L., "On the Theory of Axially Symmetric Electron Beams in an Axial Magnetic Field," Proc. IRE, Vol. 37, No. 10, pp. 1252-1258, Nov. 1949
10. Susskind, C., "A Convergent Hollow Beam Electron Gun for Microwave Tubes," Stanford Electronics Labs and Microwave Labs, Stanford Univ., Stanford, Calif., Tech. Report No. 450-1, June 29, 1956
11. Kino, G. S. and J. J. Taylor, "The Design and Performance of a Magnetron Inspection Gun," IRE Trans. on Electron Devices, Vol. ED-9, No. 1, pp. 1-11, January 1962
12. Smullin, L., Noise in Electron Devices, John Wiley, 1959
13. Glass, M.S., "Straight Field Permanent Magnets of Minimum Weight for TWT Focusing-Design and Graphic Aids in Design," Proc. IRE, Vol. 45, No. 8, pp. 1100-1105
14. Code Z1
15. Code Z2
16. Chang, K. K. N., "Confined Electron Flow in Periodic Electrostatic Fields of Very Short Periods," Proc. IRE, Vol. 45, No. 1, pp. 66-73, Jan. 1957
17. Chang, K. K. N, "Biperiodic Electrostatic Focusing for High-Density Electron Beams," Proc. IRE, Vol. 45, No. 11, pp. 1522-1527 ·
18. Siekanowitz, W.W. and F.E. Vaccaro, "Periodic Electrostatic Focusing of Laminar Parallel-Flow Electron Beams," Proc. IRE, Vol. 47, No. 3, pp. 451-452, March 1959
19. Belohoubek, E. F., "Slow-Wave Structures for Electrostatically-Focused High-Power Traveling-Wave Tubes," RCA Review, Vol. 21, No. 3, pp. 377-388, Sept. 1960
20. Cook, J. S., R. Kompfner, and C. F. Quate, "Coupled Helices," Bell Syst. Tech. Jour., Vol. 35, No. 1, pp. 127-178, January 1956
21. Kompfner, R., "Traveling Wave Valve-New Amplifier for Centimetric Wavelengths," Wireless World, Vol. 52, pp. 369-372, Nov. 1946
22. Based on work done by J. R. Pierce and L. M. Field
23. Laico, J.P., H.L. McDowell, and C. R. Master, "A Medium-Power Traveling-Wave Tube for 6,000 Mc Radio Relay," Bell Syst. Tech. Jour., Vol. 35, No. 6, pp. 1285-1346, Nov. 1956
24. Wolkstein, H. J., "Effect of Collector Potential on the Efficiency of Traveling-Wave Tubes," RCA Review, Vol. 19, No. 2, June 1958
25. Sterzer, F., "Improvement of Traveling-Wave Tube Efficiency through Collector Potential Depression," IRE Trans. on Electron Devices, Vol. ED-5, No. 4, October 1958

Article Outlines

Emitters and Photoconductors

Fundamentals of Electron Emission

L. S. Nergaard

ELEMENTARY CONSIDERATIONS..... 1
- The Dual Nature of Particles
- Quantization of Energy
- Quantum Statistics
  - Fermi-Dirac Statistics
  - Bose-Einstein Statistics

ELEMENTARY SOLID-STATE PHYSICS..... 2
- Some Properties of Solids
- Electronic Properties of Solids
  - Band Structure
  - Conductivity
  - Diffusion
  - Donors, Acceptors, and Traps
- Conclusion

EMISSION MECHANISMS..... 8
- Thermionic Emission
- Field Emission
- Photoemission
- Secondary Emission

THE PHYSICS OF THE OXIDE CATHODE 15
- Emission-Current Density
- Model of the Oxide Cathode
- Cathode Conductivity
- Pulse Operation of Oxide Cathodes
- Donors and Their Formation

APPENDIX A. THE "WAVE EQUATION" FOR ELECTRONS, AND THE DENSITY OF STATES 23
- The Wave Equation
- A Simple Case of Quantization
- The Density of States for a Free Electron

APPENDIX B. FERMI-DIRAC STATISTICS..... 25
- The Transitions Between Two "States"
- Transitions to "Free" Energy States

LIST OF SYMBOLS..... 26

REFERENCES..... 27

Oxide-Coated Emitters

C. P. Hadley

GENERAL DESCRIPTION..... 28

THE CATHODE COATING..... 28

THE BASE MATERIAL 29

DECOMPOSITION OF BINDER AND CARBONATES..... 30

ACTIVATION..... 30

MECHANISM OF OPERATION 31

THE MEASUREMENT OF THERMIONIC EMISSION..... 32

DYNAMIC EQUILIBRIUM AND POISONING..... 35
- Activating Mechanisms
- De-Activating Mechanisms
- Destruction of Barium, Strontium, or Calcium Oxides

CONCLUSIONS 35

LIST OF SYMBOLS..... 35

REFERENCES 36

Electron-Emission Coating for the Oxide Cathode

C. H. Meltzer and E. G. Widell

BASIC CONSIDERATIONS FOR THE EMISSION COATING 37
- Work Function
- Semiconductor Plus "Electron-Pore Gas" Conductivity
- Activation of the Oxide-Cathode System
- Electrical Importance of the Substrate Metal and Emission Coating
- Equilibrium Systems and Interactions

PREPARATION OF THE EMISSION COATING..... 40
- The Carbonate Form
- The Oxide Form

THERMODYNAMIC CONSIDERATIONS OF OXIDE-CATHODE SYSTEMS..... 50

DECOMPOSITION OF THE EMISSION COATING 52

AMMONIUM-PRECIPITATED CARBONATES 54

LIFE ASPECTS OF THE OXIDE-COATED CATHODE.... 56
- The Manufacturing Operation
- Application End Use

CONCLUSION..... 61

REFERENCES..... 61

Nickel Base Metal for the Oxide Cathode

C. H. Meltzer and E. G. Widell

NICKEL 64

REQUIREMENTS OF NICKEL ALLOYS FOR CATHODE BASE MATERIAL..... 64

METALLURGY OF NICKEL AND ITS ALLOYS..... 65
- Air-Melt Process
- Vacuum-Melt Process

EFFECTS OF REDUCING AGENTS ON HOT STRENGTH AND HOT RESISTIVITY 68

EFFECTS OF REDUCING AGENTS ON CATHODE ACTIVITY AND REACTIONS..... 70
- Barium Vapor Pressure and Diffusion Rates
- Sublimation Effects

REVIEW OF SPECIFIC REDUCING AGENTS..... 72
- Carbon in Pure Nickel—N81
- Carburized N18 Cathodes
- Carbon and Magnesium
- Magnesium
- Silicon
- Aluminum
- Tungsten
- Titanium
- Manganese
- Calcium
- Boron

CHARACTERISTICS OF SELECTED CATHODE ALLOYS.. 75

REFERENCES..... 75

The Interface and Other Oxide-Cathode Impedances

E. J. Hannig

CATHODE SLEEVE ACTIVITY 77
CATHODE ACTIVATION 78
CATHODE IMPEDANCES 79
EQUIVALENT CIRCUIT OF CATHODE IMPEDANCES 80
DECAY CHARACTERISTICS 82
EFFECT OF ALLOY ON CATHODE IMPEDANCE 83
OPERATIONAL CAUSES OF INTERFACE IMPEDANCE ... 83
COATING ADHERENCE 84
CATHODE IMPEDANCE EVALUATION 84
- Sine-Wave Methods
- Square-Wave Methods
- Pulse Method

REFERENCES 87

Tungsten, Thoriated-Tungsten, and Thoria Emitters

W. E. Harbaugh

PURE-TUNGSTEN EMITTERS 90
- Preparation of Tungsten Filaments
- Properties of Tungsten Emitters
- Present Uses

THORIATED-TUNGSTEN EMITTERS 91
- Preparation
- Reasons for Carburization
- The Carburization Process
- End-Loss Correction
- Preferred Carburization
- Some Precautions for Carburization
- Quality Control

THORIA EMITTERS 96
SUMMARY 98
REFERENCES 98

Photoelectric and Secondary-Electron Emission

A. H. Sommer

PHOTOELECTRIC EMISSION

GENERAL CONSIDERATIONS 99
METALLIC PHOTOEMITTERS 99
SEMICONDUCTING PHOTOEMITTERS 100
PRACTICAL PHOTOCATHODES 101
- The Antimony-Cesium (Sb-Cs) Cathode
  - Formation
  - Properties
- Miscellaneous Cathodes of the Sb-Cs Type
- Multi-Alkali Cathodes
- The Silver-Oxygen-Cesium (Ag-O-Cs) Cathode
  - Formation
  - Properties
- The Bismuth-Silver-Oxygen-Cesium (Bi-Ag-O-Cs) Cathode
  - Formation
  - Properties
- Matching of Photocathode and Light Source

SECONDARY-ELECTRIC EMISSION

GENERAL CONSIDERATIONS 105
PARAMETERS OF SECONDARY-ELECTRON EMISSION .. 105
- Gain Factor δ as a Function of Primary Energy
- Energy Distribution of Secondary Electrons
- Gain Factor δ as a Function of Angle of Incidence

PRACTICAL SECONDARY EMITTERS 106
- Secondary-Emitting Photocathodes
- Silver-Magnesium-Oxygen (Ag-Mg-O)
- Copper-Beryllium-Oxygen (Cu-Be-O)

REFERENCES 106

Photoconductors for Camera Tubes

C. W. Rector

BASIC THEORY 108
APPLICATIONS 112
- Germanium
- Cadmium Sulfide
- Microcrystalline Photoconductors
- Amorphous Photoconductors

PHOTOCONDUCTORS FOR THE VIDICON 115
- Antimony Trisulfide Photosurfaces
- Effects of Pressure
- Effects of Photolayer Thickness
- Effects of the Substrate
- Effects of Modifications of the Basic Photoconductor
- Miscellaneous Evaporation Parameters

FACTORS INFLUENCING SELECTION OF PHOTOCONDUCTORS FOR THE VIDICON 122
- General Requirements of Photoconductors for the Vidicon
- Spectral Photoconductive Response

COMPOSITION GRADED PHOTOSURFACES 126
REFERENCES 127

Phosphors

Cathodoluminescent Materials

R. N. Summergrad

THE CATHODOLUMINESCENT PROCESS 128
PREPARATION OF PHOSPHORS 129
PHOSPHOR RESEARCH 131
PHOSPHOR CHARACTERISTICS 133
- Phosphor Efficiency
- Decay Characteristics of Phosphors
- The Variation of Light Output with Accelerating Potential and with Current Density
  - Accelerating Potential
  - Current Density
- Temperature: Its Effect on Luminescence

REFERENCES 136

Application and Colorimetry of Phosphor Screens

A. E. Hardy

SCREEN APPLICATION TECHNIQUES 138
APPLYING PHOSPHOR SCREENS FOR BLACK-AND-WHITE PICTURE TUBES AND OSCILLOSCOPE TUBES 138
STOKE'S LAW 139
SCREEN ADHERENCE 139
FILMING AND ALUMINIZING BLACK-AND-WHITE SCREENS 139
SCREENS FOR COLOR TUBES 140
EXPOSURE SYSTEM 141
DEVELOPING 141
ONE-STEP SLURRY PROCESS 141
SCREEN DEFECTS 143
- Cross-Contamination
- White Nonuniformity
- Blue-Phosphor Breakdown

FILMING AND ALUMINIZING OF COLOR-TUBE SCREENS 143
SCREEN BAKEOUT 144
PHOTOMETRY OF PHOSPHOR SCREENS 144
MEASURING INSTRUMENTS 145
COLOR MEASUREMENT AND SPECIFICATION 145
APPLICATION OF C.I.E. COLOR SYSTEM TO BLACK-AND-WHITE SCREENS 146
APPLICATION OF C.I.E. COLOR SYSTEM TO COLOR TELEVISION SCREENS 148
REFERENCES 151

Manufacture of Cathodoluminescent Materials

J. A. Markoski

SULFIDE PHOSPHORS 153
- Preparation of Stock Solutions
- Precipitation
- Neutralization and Washing
- Drying
- Activation and Fluxing
- Firing and Coating

OXYGEN-DOMINATED PHOSPHORS 157
- Zinc Orthosilicate: Manganese Activated
- Zinc Orthophosphate: Manganese Activated

REFERENCE 158

Space Current

Calculation of Fields and Currents

L. C. Scholz

ELECTROSTATIC FIELDS 159
SOLUTIONS FOR SIMPLE GEOMETRIES 159
- Infinite Parallel Planes
- Concentric Cylinders

THE THERMIONIC DIODE 159
THE IDEAL INFINITE PARALLEL-PLANE DIODE 161
- Initial Velocity - Zero
- Initial Velocities - Maxwellian
- Discussion

THE IDEAL CYLINDRICAL DIODE 168
- Initial Velocities - Zero
- Initial Velocity - Maxwellian
- Discussion

THE EXTENSION OF DIODE THEORY TO TRIODES 170
- The Shielding Effect of a Grid
- Calculation of Equivalent Potentials
- Interpretation
- Procedure for Analysis

TETRODES AND PENTODES 180
- Equivalent Potentials
- The Effect of Space Charge in the Anode Region
  - Qualitative Discussion
  - Mathematical Solution-References
  - Screen-Grid Current
  - Plate Resistance

VARIABLE MU EFFECTS 182
- The Concept of Variable Mu
- Design of Variable-Mu Tubes
- Mu Variations from Nonideal Grids
- Variable Spacing and Statistical Variations

FIELD PLOTTING 184
- Mathematical Plotting Methods
  - Direct Solution of the LaPlace Equation
  - Conformal Mapping
  - Numerical Solutions
- Analog Plotting Methods
  - The Resistance-Paper or Electrolytic-Tank Method
  - The Resistor Net
- Field Sketching
- Electrode Fitting

TRAJECTORY TRACING 189
- General Equations of Motion
- General Projectory in Two Dimensions
- Electron Motion in a Uniform Electric Field
- Snell's Law
  - Segmented Straight-Line Trajectories
  - Mechanics of Trajectory Tracing
  - Segmented Parabolic Trajectories
- Discussion
- Nomographs for the Rapid Modification of Electron-Tube Characteristics
- Theoretical Basis for Nomographs
- Applications and Limitations of the Nomographs
- Use of Multiplying Factors on Nomographs

REFERENCES 198

Space Current Flow

M. V. Hoover

THE ELECTRON 202
EMISSION OF ELECTRONS 202
- Thermionic, or Primary, Emission
- Secondary Emission
- Photoelectric Emission
- High-Field Emission
- Radioactive Disintegration

LAWS OF ELECTRONS IN MOTION 206
- Electron Motion in a Uniform Electric Field
- Electron Motion in a Uniform Magnetic Field
- Electron Motion in Combined Electric and Magnetic Fields

ELECTRON OPTICS 208
Electrostatic Electron Optics
Magnetic Lenses
Periodic Focusing
Electron Guns
ELECTRON SPACE CHARGE 210
Space-Charge Effects in Diodes
Space-Charge Effects in Triodes
Effect of Space Charge on Electron Transit-Time
Space Charge and the "Virtual" Cathode
Space Charge and Electron-Beam Dispersion
Space-Charge Waves
INTERACTIONS OF ELECTRONS WITH CIRCUITS.... 213
Exchange of Energy Between Electron Streams and Electric Fields
Circuit Currents Induced by Electron Motion
Electron-to-Circuit Energy-Exchange Systems
Electron Beams in a Nonpropagating-Circuit Environment
REFERENCES..... 214

General Design Considerations for Receiving Tubes

Heaters

A. G. F. Dingwall

BASIC DESIGN CONSIDERATIONS AND THEORY 216
Heater Wire Dimensions
Heater Design Temperature
Insulation Thickness
Apex Chip Size
Heater Warm-Up Time
HEATER TYPES 224
Folded Heaters
Double-Helical Heaters
Folded Single-Helical (Hairpin) Heaters
HEATER MATERIALS..... 225
Heater Wire Materials
Tungsten as a Heater Material
Dowmo as a Heater Material
Insulation Coating Materials for Heaters
APPLICATION OF HEATER COATINGS..... 228
HEATER BEHAVIOR ON LIFE..... 229
Failure-Rate Patterns and Heater Temperature
Heater-Cathode Bias
Heater-Positive Biases
Heater-Negative Biases
HEATER-CATHODE LEAKAGE 232
Definition and Characteristics
Decay and Polarization
Distribution of Heater-Cathode Leakage Values
Changes of Leakage after Tapping of Tube
Effect of Heater Temperature on Leakage
The Effects of Heater Size
The Mechanisms of Heater-Cathode Leakage
Factors Contributing to Heater-Cathode Leakage
Heater Hum
Aging
REFERENCES..... 242

Heat Transfer in Receiving Tubes

O. H. Schade, Jr.

HEAT TRANSFER BY STEADY, UNIDIRECTIONAL CONDUCTION 244
The Fourier Equation and Two Simple Solutions
The Concept of Thermal Resistance
HEAT TRANSFER BY CONVECTION 245
THE PHENOMENON OF HEAT TRANSFER BY RADIATION THROUGH NON-ABSORBING MEDIA..... 245
The Physics of Radiation
The Black-Body Concept
The Distribution of Radiation
The Gray-Body Concept
The Use of Spectral and Total Emissivities
SOLUTIONS OF HEAT TRANSFER EQUATION FOR RADIATION THROUGH NONABSORBING MEDIA..... 247
Black-Body Radiation for Simple Geometries
The View Factor
Radiation from "Composite Surfaces"
Radiation from a "Right-Angled" Surface at Constant Temperature
A U-Shaped Radiating Surface at Constant Temperature
THE APPLICATION OF THEORY TO HEAT TRANSFER PROBLEMS IN TUBE STRUCTURES..... 253
Heat Transfer Circuits of the Diode
Heat Transfer Circuits of the Triode
Heat Transfer by Conduction Through the Grid and Stem Leads
Stem Lead to Each Siderod
Stem Lead to Only One Siderod
Heat-Transfer Considerations for Multigrid Tubes
EXAMPLE OF HEAT TRANSFER CALCULATIONS—THE 6L6G PLATE AND SCREEN-GRID TEMPERATURES..... 261
Solution for the Average Plate Temperature
Solution for the Average Screen-Grid Temperature
LIST OF SYMBOLS 262
REFERENCES..... 264

Contact Potential and Grid Currents

E. R. Schrader

WORK FUNCTION AND CONTACT POTENTIAL
WORK FUNCTION..... 266
CONTACT POTENTIAL 266
THREE METHODS OF MEASURING CONTACT POTENTIALS.. 267
The Intersection Method for Measuring Contact Difference of Potential (Method I)
Significance of the Contact Potential Curve
The Method of Measuring a Potential to Produce an Arbitrary Reference Current (Method II)
The Method of Measuring Potential with a Grid Voltmeter (Method III)

GRID CURRENTS

CATHODE-TO-GRID EMISSION 273
PRIMARY GRID EMISSION 274
LEAKAGE CURRENTS 274
GAS ION CURRENTS 274
THE COMPOSITE GRID CHARACTERISTIC 275
INTERPRETATION OF DATA 275
REFERENCES 276

A New Approach to the Calculation of Electron Tube Characteristics

O. H. Schade, Sr.

CLASSICAL METHOD OF ELECTRON-TUBE CALCULATION

THE PROBLEM OF ELECTRON-TUBE CALCULATION 278
CHARACTERISTICS OF THE "PERFECT" DIODE 279
- Work Functions (ψ), Electron Velocities (V_T), and Saturation Current (I_S)
- The Diode Circuit (Potentials and Operating Conditions)
  - The Saturation Region
  - The Space-Charge-Limited Region
  - The Exponential Point
  - The Exponential Region
  - Effective Anode Potential V and Reference Condition (V = 0)
- The Langmuir Relations
- Calculation of the Characteristic I = f(V)
- Calculation of the Characteristic I = f (E_b)
- Some Properties of the Perfect Diode
  - The Barrier Potential ($\psi_c + V_m$)
  - Exponential Region V < 0
  - Positive-Accelerating-Potential Region (V > 0)
  - Diode Conductance g = f(T, ψ_c, d)

TRIODES 287
- Electrostatic Capacitances and Potentials in "Perfect" Parallel-Plane Triodes (μ = Constant)
- Correction for the Space Charge by an Equivalent Diode Spacing (d_1')
- Aberrations in Practical Triodes and Limitations of the Simple Theory
- The Grid at Space Potential
- The Grid Work Function ψ_g
- Cathode-Temperature Variations $+\Delta T$
- Eccentric Electrodes
  - Uniform Eccentricity of Grid and Cathode
  - Eccentricity Caused by Grid Ellipticity
  - Grid-Plate Eccentricity Δd_2
  - Nonuniform Eccentricity (Tilting) of Cylindrical Electrodes
- Calculation of Vibration-Induced Modulation in Cylindrical Structures

A REFINED METHOD FOR CALCULATING THE SPACE CURRENT DISTRIBUTION AND CURRENT-VOLTAGE CHARACTERISTICS OF ELECTRON TUBES

A NEW APPROACH TO SPACE-CURRENT CALCULATION IN ELECTRON TUBES 295
- The Electrostatic Field in the Cathode-Grid Space
- Amplification Factors (μ)
- Effective Potentials and Space-Current Distribution

THE COMPUTER PROGRAM 300

A. THE MACHINE PROGRAM FOR PARALLEL-PLANE TRIODES 300
- Calculation of Electrostatic Potential Tables (Sub-Program)
- Calculation of (V,I) Characteristics (Main Program)
- Computed Data
  - Computed Potentials (V_g, V_a) and External Voltages (E_c^g, E_b)
  - The Derivatives (gm, r_p, $\bar{\mu}$)
- Input Information
  - Determination of Potential Scales and Increments ΔV_g and ΔV_a
  - Starting Values for V_a and V_g and Current Cutoff

B. ALTERNATE PATHS AND SUBROUTINES EXTENDING FLEXIBILITY AND SCOPE OF THE PROGRAM 307
- Subroutine for a Direct Comparison of Computed Characteristics
- Subroutines for Triodes Having Cylindrical Elements
- Correction for Mesh Grids
- Nonuniform Spacings or Parameters (Segmentation Routine)

CORRELATION OF COMPUTED AND MEASURED DATA .. 311
- Measurement of (V,I) Characteristics
- Analysis of Measured Characteristics

CONCLUSIONS 313
ACKNOWLEDGMENTS 314
REFERENCES 314
LIST OF SYMBOLS 314

Metals, Metallurgy, and Metal Fabrication

Metals and Metallurgy

C. W. Horsting, et al

THE METALLIC STATE
C. W. Horsting

ATOMIC STRUCTURE 315
GRAIN STRUCTURE 316
ALLOTROPY 317
DUCTILE DEFORMATION 317
DISLOCATIONS 318
TENSILE TESTING 319
ANNEALING AND RECRYSTALLIZATION 319
DIFFUSION 320
PHASE EQUILIBRIA IN ALLOY SYSTEMS 320
PHYSICAL PROPERTIES OF ALLOYS 322
HARDENING TREATMENTS 322
VAPOR PRESSURE OF METALS 323
REFERENCES 323

METALS FOR ELECTRON TUBES
J. J. Carrona

NICKEL 324
IRON AND STEEL 326
COPPER 326
TUNGSTEN 327
MOLYBDENUM 328

OTHER METALS 329
Silver
Aluminum
Chromium
Tantalum
Titanium and Zirconium
REFERENCES 330

TESTING OF METALS
C. H. Li

MECHANICAL TESTING 330
Tensile Tests
Hardness Tests
Stiffness Test
Springback Test
Ductility Test
METALLOGRAPHIC TESTING 331
X-RAY TESTING 331
Industrial Radiography
Diffraction Study
THERMAL TESTING 332
Dilatometric Analysis
Emissivity Measurements
Thermal Analysis
CHEMICAL TESTING 332
Compositions Analysis by Chemical and Spectrographic Means
Rapid Identification of Metals
Corrosion Tests
ELECTRICAL TESTING 332
Resistivity Checks
Burnout Tests
MISCELLANEOUS TESTING 333
Weighing
Powder Examination
Surface Test for Identifying Metals
Spark Testing
Color Comparison Test

CATHODE MATERIALS FOR RECEIVING TUBES
I. S. Solet

DIRECTLY HEATED CATHODES 333
INDIRECTLY HEATED CATHODES 334
GASES IN NICKEL 334
"IMPURITIES" IN NICKEL AND THEIR RELATION TO EMISSION 334
THE FUTURE OF THE THERMIONIC CATHODE 335
REFERENCES 336

PLATE MATERIALS (METALS)
H. G. Scheible and T. A. Sternberg

PROPERTIES OF COMMON STRIP MATERIALS 336
Clear Nickel, Nickel-Plated Steel, and Nickel-Clad Steel
Carbonized Nickel, Nickel-Plated Steel, and Nickel-Clad Steel
GAS-CARBONIZED MATERIALS 337
SLURRY-CARBONIZED MATERIALS 337
ALUMINUM-CLAD STEEL 337
ALNIFER 337
COPPER-BASE AND COPPER-CORED MATERIALS 338
PROPERTIES OF SOME TYPICAL PLATE MATERIALS 338
BIBLIOGRAPHY 339

GRIDS AND LEADS
A. J. Stoeckert

CONSTRUCTION 339
Materials—Lateral Wires
Materials—Side Rods
LEADS 341
LIST OF MANUFACTURERS OF ALLOYS 344

METAL-TO-METAL JOINING
E. S. Thall

ELECTRIC WELDING 344
BRAZING AND SOLDERING 346
COLD WELDING 348
RIVETING, FOLDING, AND NOTCHING AND PEENING 349
REFERENCE 349

JOINING METALS TO GLASSES AND CERAMICS
K. Strater

METAL-TO-GLASS SEALS 350
METAL-TO-CERAMIC SEALS 354
REFERENCES 356

SYSTEMATIC CHARTS OF THE PROPERTIES OF ELECTRON TUBE METALS
S. Y. Husni

This section contains tabular and graphic information on some of the properties of metals and alloys that should be of interest to the electron tube engineer. The following tables and charts are included:

Properties of Metals (Table XVIII) 358-359
Properties of Alloys (Table XIX) 360-363
Coefficients of Thermal Expansion at Elevated Temperatures (Table XX) 364
Tensile Strengths at Elevated Temperatures (Table XXI) 364
Total Emissivity (Table XXII) 365
Thermal Conductivity at Elevated Temperatures (Table XXIII) 365
Charts of Vapor Pressures at Elevated Temperatures (Figs. 32, 33) 366-369

Metals for Vacuum Tube Construction

P. D. Strubhar

COPPER 370
GLASS-SEALING METALS AND ALLOYS 371
TUNGSTEN AND MOLYBDENUM 373
STEEL 375
OTHER METALS 376

Principles of Metal Forming

L. S. Sloane

THE MECHANICAL PRESS 377
PRESS OPERATIONS 377
BLANKING AND PIERCING 377

DRAWING 378
Blankholding Pressure
Draw Radius
Material Reduction
Diameter Reduction
Blank Diameter
Reverse Drawing
SIZING 381
COINING 381
EXTRUDING 381
FORMING 381
TOOLING 382
Tool Costs
Component Costs
FOUR-SLIDE MACHINES 382

Processes Used in Metal Parts Fabrication and Finishing

H. A. Kauffman

RESISTANCE WELDING OF ELECTRON TUBE PARTS AND ASSEMBLIES 385
Spot Welding
Projection Welding
Seam Welding
ELECTROPOLISHING 387
CHEMICAL BRIGHT DIPPING 387
PHOTOMECHANICAL ETCHING 387
Sensitizing
Printing
Developing
Etching
ELECTROPLATING 389
Effect of Parts Design
Requirements for Nonplated Areas
Requirements for No Nonplated Areas
Cleaning for Plating
Striking
Plating Requirements and Costs
Treatment after Plating
Inspection of Plating
Summary
REFERENCES 391

Insulators, Envelope Materials, and Seals

Glass for Receiving Tubes

J. Gallup

DEFINITION 392
COMPOSITION 392
PRINCIPAL PROPERTIES OF GLASS 393
Thermal Expansion
Viscosity
Strain Point
Annealing Point
Softening Point
Working Point
Annealing of Glass
Electrical Resistivity
PHYSICAL PROPERTIES OF LESSER INTEREST FOR RECEIVING TUBE MANUFACTURE 397
Density
Weathering Resistance
Dielectric Constant and Power Factor
OPTICAL PROPERTIES — COLOR, LIGHT TRANSMISSION, LUMINESCENCE, INDEX OF REFRACTION, AND BIREFRINGENCE 400
Color
Light Transmission, Luminescence, and Index of Refraction
Birefringence under Strain
STRENGTH, CRACKS, AND STRAIN CONTROL 402
THERMAL AND MECHANICAL TESTS 403
CHEMICAL TREATMENT OF GLASS SURFACES 404
SOME GENERAL CONSIDERATIONS CONCERNING GLASS-STEM, BULB, AND SEAL DESIGN AND USE 404
REFERENCES 405
Books
Periodicals

Glass and Its Properties and Seals

J. C. Turnbull and G. E. Eiwen

COMPOSITION AND PROPERTIES 406
Compositions of Commercial Glasses
Properties of Commercial Glasses
Description and Application of Glasses
Dumet-Sealing Glasses
Chromium-Iron Alloy Sealing Glasses
Iron-Sealing Glasses
Kovar-Sealing Glasses
Tungsten-Sealing Glasses
MECHANICAL STRENGTH OF GLASS 411
Mechanical Strength and Fatigue of Glass
Glass Fracture Analysis
Measurement of Stresses in Glass
Polariscopic Estimation of Stress
Measurement of Strains in Glass Surfaces
Scratch Testing of Stresses in Glass
Estimation of Stresses by Strength Testing of Glass Articles
ELECTRICAL PROPERTIES OF GLASS 416
Electrical Failure of Glass
Electrolysis Effects
Puncture
Ionization in Air Lines and Bubbles
Conductive Films on Glass Surfaces
OPTICAL PROPERTIES OF GLASS 418
Transmission of Light and Radiation
X-Ray Absorption
Glass Browning
Fluorescence of Glass
Homogeneity—Glasses Melted for Optical Use
OUTGASSING, DIFFUSION, AND SUBLIMATION OF GLASS 420

Thermal Outgassing of Glass
Diffusion in Glass
Diffusion of Water in Glass
Diffusion of Gases in Glass
Diffusion of Ions in Glass
Gases Dissolved in Glass
Outgassing under Bombardment
Sublimation from Glass Surfaces
DEVITRIFICATION OF GLASS 423
THERMAL EXPANSION OF GLASS 423
GLASS-TO-METAL AND GLASS-TO-GLASS SEALS 424
Nature of Sealing Problems
Adherence of Glass Sealed to Metal
Estimating Quality of Seals
Commonly Used Glass-to-Metal Seal Combinations
Sealing Methods
Stresses in Seals
Calculation of Stresses from Thermal-Expansion Differentials
Measurement of Stresses in Seals
Annealing of Seals
Practical Annealing Schedules
REFERENCES 432

Ceramics

W. F. Lawrence

WHEN TO USE CERAMICS 433
THE INTERRELATIONSHIP OF FORMING METHODS AND PART DESIGN 434
Forming Techniques Used to Make Finished Parts
Forming Techniques Suitable for Subsequent Machining
Pressing Method
Extrusion Method
TOLERANCES 438
CHOICE OF MATERIAL 440
MECHANICAL PROPERTIES 441
Strength
Impact Strength
Specific Gravity, Porosity, and Water Absorption
THERMAL PROPERTIES 442
Maximum Operating Temperature
Thermal Conductivity
Specific Heat
Thermal Expansion
Resistance to Thermal Shock
ELECTRICAL PROPERTIES 449
Dielectric Constant and Loss Factor
Resistivity
Dielectric Strength
REFERENCES 451

Ceramics and Ceramic-to-Metal Seals

M. Berg

CERAMICS FOR ELECTRONIC APPLICATIONS

TYPES OF CERAMICS FOR ELECTRON-TUBE COMPONENTS 452
Alumina
Steatites - Clinoenstatite Bodies
Forsterite
Electrical Porcelain
Zircon Porcelain
Beryllia
High-Dielectric Constant Ceramics
Other Ceramics
APPLICATIONS AND PHYSICAL PROPERTIES 455
Design and Processing
Fabrication Techniques

CERAMIC-TO-METAL SEALS

PROCESSES USED TO JOIN METAL TO THE CERAMIC ENVELOPE 458
The Sintered Metal Process
The Active Alloy Process
The Pressure Process
The Glass Interlayer Process
SEAL DESIGN 463
BIBLIOGRAPHY 464

Design of Mica Spacers

M. Bondy and M. Jinetopulos

MINING AND PROCESSING OF MICA 465
PROPERTIES OF MUSCOVITE 466
GRADING OF MICA 467
TOOLING 467
SPACER DESIGN 468
Mica Thickness
Bridge Distances
Minimum Hole Dimensions
Outside Contour
MICA COATINGS 469
VOLTAGE GRADIENTS 470
REFERENCES 471

Parts Cleaning, Coating, and Basing

Parts Cleaning and Processing

F. I. Scott

IMPORTANCE AND SCOPE OF CLEANING 472
Background
General Considerations
Classification of Contaminants
Evolution of Gases from Parts
Tests for Cleanliness
Methods of Cleaning
PROCESS AND MATERIAL CONTROLS RELATED TO CLEANING 478
Raw Materials
Lubricants
Storage and Handling
SUMMARY 480

ADDITIONAL CLEANING PROCESSES 480
Alkaline Electrocleaning
Emulsion Cleaning
Tumbling
REFERENCES 480

Coatings, Bases, and Basing Cements

L. P. Fox

INTERNAL COATINGS 481
EXTERNAL COATINGS 482
METAL COMPONENT COATINGS 482
PROTECTION OF SILVER-PLATED TUBES 483
BASES 483
BASING CEMENTS 483
REQUIREMENTS FOR GOOD BASE ADHERENCE 483

Exhaust and Finishing

High-Vacuum Technology and Equipment

W. G. Henderson and A. L. Lucarelli

WHAT IS A VACUUM? 485
PUMP RANGE 485
VACUUM CLASSIFICATION 485
TYPES OF FLOW AND MEAN FREE PATH 485
SYSTEM COMPONENTS 486
GETTERING 486
REQUIREMENTS FOR A GOOD VACUUM 487
MECHANICAL PUMPS 487
MECHANICAL-PUMP SPEEDS 488
EXHAUST TIME - MECHANICAL PUMP 488
GAS BALLAST PUMP 489
MECHANICAL BOOSTER PUMP 490
OIL-DIFFUSION PUMP 491
FRACTIONATING PUMP 491
BACKSTREAMING 491
PURPOSE OF CONSTRUCTION 492
PUMP COOLING 492
PUMP CURVE FEATURES 492
SYSTEM CURVES 493
VAPOR PRESSURE - PUMP MEDIUM 494
VAPOR PRESSURE - METALS 495
VAPOR PRESSURE - GASKET MATERIALS 495
EVAPOR-ION PUMP 495
MERCURY PUMP 496
GAUGES 497
SEALS 500
TRAPS AND BAFFLES 501
VALVES 502
FLOW OF GASES THROUGH VACUUM LINES 503
STARTING AND STOPPING PROCEDURE 503
OTHER OPERATING AND DESIGN PRECAUTIONS 503
TROUBLE SHOOTING 503
OTHER PROCESSES 504
Sputtering
Vacuum Firing — AC Heating
Vacuum Firing — RF Induction Heating
Vacuum Firing — Electric Oven
REFERENCES 505
BIBLIOGRAPHY 505

Exhaust of Receiving-Type Tubes

C. J. Pearce

OBJECTIVES OF THE EXHAUST PROCESS 506
CATHODE BREAKDOWN AND BRIGHTNESS TEMPERATURE 506
TRACK LIGHTING SEQUENCE 506
Temperature - Time - Material Effects
Pressure - Time Relationships
PARTS OUTGASSING AND COLOR TEMPERATURES 508
Temperature - Time - Material Effects
Parts Processing History
Simultaneous Parts Degassing
Interactions of Cathode Breakdown and Parts Outgassing
VACUUM GAUGES 510
Absolute Gauges
Relative Gauges
FORMS OF RECEIVING-TYPE EXHAUST EQUIPMENT 513
Manual Exhaust (Trolley)
Manual Exhaust Schedule for Vacuum-Type Tubes
Manual Exhaust for Gas-Filled Tubes
Automatic Exhaust (Sealex)
Sealex-Exhaust Schedules for Glass-Type Vacuum Tubes
Sealex Exhaust of Metal-Type Vacuum Tubes
Sealex Exhaust of Gas Tubes
Effects of Exhaust Tubulation Size
Other Methods of Exhaust
REFERENCES 517
BIBLIOGRAPHY 517

Getters

C. H. Thomas

GETTER ACTION 519
HISTORICAL TYPES OF GETTERS 519
Phosphorus
Magnesium
Barium Azide
Misch Metal
Barium
Barium-Aluminum-Magnesium Pellet
Barium-Aluminum Pellet
Batalum Getter
Kic Getter

King Getter
Ring-Type Getter
Zirconium and Titanium
Tantalum
Cer-Alloy
Cerium-Thorium Alloy
EFFICIENCY OF GETTERS 522
EFFICIENCY OF GETTER FLASH..... 524
Preflash
Postflash
Flashing in Argon
DETERMINATION OF BARIUM IN GETTER FLASH.... 524
REFERENCES..... 525

Aging and Stabilization

G. Wolfe

AGING

BASIC CONSIDERATIONS..... 526
THE AGING PROCESS..... 527
The Hotshot Step
The Grid Step
The Anode (or Plate) Step
General Rules of Procedure
Summary
CIRCUITRY AND EQUIPMENT..... 529
Basic Circuits
VARIATIONS OF TREATMENTS AND SPECIAL PROCESSING..... 532
Elimination of Grid Step
Special Cycling
Problem Tube Types
Contact Potential Control
Parasitics

STABILIZATION

AREAS OF APPLICATION..... 533
Passive-Type Cathodes
Elimination of Potential Inoperatives
CIRCUITS AND EQUIPMENT 534

SUMMARY

FACTORS AFFECTING AGING AND STABILIZATION SCHEDULES 535
Relation of Aging to Exhaust
Tube Parts and Materials
Tube Design
IMPORTANCE OF MEASUREMENTS..... 535
BIBLIOGRAPHY 536

Design, Processing, and Application of Receiving-Type Tubes and Picture Tubes

Diodes

A. P. Kauzmann

DIODE SYMBOLS..... 537
DIODE THEORY 537
Plane Parallel Electrodes
Concentric Electrodes
Perveance
Diode-Plate Characteristic
Dynamic and Effective Diode Resistances
Diode with Added Series Resistance
DIODE OPERATION INTO A SHORT-CIRCUITED LOAD.... 542
BASIC RECTIFIER POWER CONVERTERS 542
RESISTANCE LOAD AND CHOKE-INPUT-FILTERED LOAD.. 543
TUNGSTEN-LAMP-LOADED RECTIFIER..... 548
CAPACITOR-INPUT-FILTERED LOAD..... 549
CAPACITOR-INPUT-FILTER CIRCUITS USING LARGE VALUES OF CAPACITANCE..... 552
VOLTAGE-MULTIPLYING CIRCUITS..... 569
HIGH-VOLTAGE RECTIFIER TUBE DESIGN..... 571
Potential Gradients and Field Emission
Secondary Emission
Potential Gradient at the Cathode
Electrolysis
X-Rays
Diode Currents
DAMPER-TUBE RECTIFIER DESIGN..... 574
REFERENCES..... 576

Small-Signal Radio-Frequency Amplifier Tubes

W. A. Harris

TELEVISION INPUT-STAGE REQUIREMENTS..... 577
MEASUREMENTS OF OPERATING CHARACTERISTICS... 577
DEFINITIONS AND THEORY..... 578
Measurements of a Two-Port
Three-Terminal Transducers
Measurement of Noise
Corrections for Lead Inductances and Capacitances
LARGE-SIGNAL CONSIDERATIONS..... 592
TUBE DESIGN CONSIDERATIONS..... 597

Design of Intermediate-Frequency Amplifier Tubes

M. B. Knight

GRID-TO-PLATE TRANSCONDUCTANCE..... 599
INPUT ADMITTANCE..... 600
OUTPUT ADMITTANCE..... 601
FEEDBACK CAPACITANCE 602
GAIN CONTROL..... 604
DISTORTION..... 604
REFERENCES..... 606

Design of Converter and Mixer Tubes

T. J. Henry

CHARACTERISTICS OF CONVERTER AND MIXER TUBES. 607
THE DESIGN OF THE CONVERTER STAGE 610
- Automatic Gain Control (AGC)
- Tuning
- Interaction of Signal and Oscillator Circuits
- Oscillator Strength
- Input Conductance at High Frequencies
- Noise
- Oscillator Radiation
- Frequency Drift During Warm-up of Television Receivers
- Economic Considerations

TYPES OF MIXERS AND CONVERTERS 612
- The Diode Mixer
- The Triode Mixer
- The Pentode Mixer
- The Outer-Grid-Injection Mixer
- The Inner-Grid-Injection Converter

COUPLING EFFECTS 615
NOISE IN MIXER TUBES 616
DESIGN OF MIXER TUBES 617
- Triode Mixers
- Pentode Mixers
- Outer-Grid-Injection Mixers
- Inner-Grid-Injection Converters
- Application Testing

CURRENT CONTROL BY A CONTROL GRID FOLLOWING A POSITIVE SCREEN GRID 618
REFERENCES 620

Design of AM and FM Detector Tubes

T. J. Henry

DETECTORS 622
DETECTION OF AMPLITUDE-MODULATED SIGNALS 622
- General Considerations
- The Diode Envelope Detector
- Other Envelope Detectors
- Single-Sideband and Vestigial-Sideband Modulation
- Product Detectors
- Noise Advantage of Product Detectors
- The Design of Tubes for Amplitude-Modulation Detector Circuits

DETECTION OF FREQUENCY-MODULATED SIGNALS 629
- General Considerations
- Types of Frequency-Modulation Detection Systems
- Detection by Amplitude Variations
- Detection by Phase Variation with Frequency

REFERENCES 635

Vacuum-Tube Oscillators

W. E. Babcock

OPERATION AND ANALYSIS OF TYPICAL OSCILLATORS . 637
- Oscillator Operation
- Sinusoidal Oscillators
- Oscillator Analysis by Feedback Amplifier Theory
- Negative-Resistance Oscillators
- Feedback Oscillators
- Electron-Coupled Oscillator
- Grid-Circuit Considerations

OSCILLATOR CIRCUITS FOR SPECIAL APPLICATIONS. 644
- Crystal Oscillators
- Magnetostriction Oscillator
- Low-Frequency Oscillators

NONSINUSOIDAL OSCILLATOR CIRCUITS 649
- Multivibrator Circuits
- Sawtooth Generators

TUBES FOR USE IN OSCILLATOR CIRCUITS 654
- Supply-Voltage Considerations
- Power Output Considerations
- Frequency Considerations
- Sinusoidal and Nonsinusoidal Oscillator Considerations

REFERENCES 655

Beam Power Tube Design Considerations

O. H. Schade, Jr.

THE BEAM POWER TUBE IN CROSS SECTION 656
SPACE-CHARGE EFFECTS 657
THE ELECTRON-BEAM CHARACTERISTICS 660
CALCULATION OF THE APPROXIMATE BEAM SHAPE AT ZERO CONTROL-GRID POTENTIAL 663
DIRECTLY INTERCEPTED SCREEN-GRID CURRENT 664
HIGH VOLTAGE CONSIDERATIONS 664
- Some Observations on the Breakdown Phenomena
- Considerations for Breakdown between Metal Electrodes
- Considerations for Breakdown along Insulator Surfaces

THE MECHANICAL-THERMAL-ELECTRICAL DESIGN OF BEAM TUBES 667
- Design Procedure for the "Optimum" Tube
- Some General Characteristics of Tube Structures

LIST OF SYMBOLS 668
REFERENCES 668

Audio Output, Vertical-Deflection, and Horizontal-Deflection Tubes for Receiving Applications

M. Bondy

AUDIO OUTPUT TUBES

DESIGN OBJECTIVES 669
- Power Output
- Power Sensitivity
- Distortion
- Plate-Circuit Efficiency

DESIGN CONSIDERATIONS 669
- Triodes
- Pentodes
- Beam Power Tubes

DESIGN EXAMPLE 671
Objectives of Design Problem
Analysis of Design Problem
Possible Solutions
Critical Design Areas

VERTICAL-DEFLECTION TUBES

DESIGN OBJECTIVES 675
DESIGN CONSIDERATIONS 676
Triodes
Beam Power Tubes
DESIGN EXAMPLE 676
Choice of Tube
Critical Design Areas

HORIZONTAL-DEFLECTION TUBES

DESIGN OBJECTIVES 678
DESIGN EXAMPLE 679
DESIGN CONSIDERATIONS 679

Black-and-White Picture Tubes

C. W. Thierfelder, et al

ELECTRON GUN DESIGN CONSIDERATIONS

DESCRIPTION AND OPERATION 682
ELECTRON LENS THEORY 683
DEFLECTION 683
TRIODE AND TETRODE CHARACTERISTICS 684
CUTOFF VOLTAGE 685
Design Parameters
Mechanical Variations
Practical Manufacturing Techniques
PREFOCUSING LENS 686
ION TRAP LENS 688
FOCUSING LENSES 690
Magnetic Lens
Electrostatic Acceleration Lens
Unipotential Electrostatic Lens
LENS ERRORS 691

PICTURE TUBE PERFORMANCE CHARACTERISTICS

ELECTRON-GUN REQUIREMENTS AND FOCUS CHARACTERISTICS 691
PICTURE QUALITY 693
Subjective Evaluation of Picture Quality
Factors Determining Picture Quality
Objective Measurement of Picture Quality
RELIABILITY 695
TRANSFER CHARACTERISTICS 696
GUN EFFICIENCY 697

TUBE DESIGN CONSIDERATIONS

BULB DESIGN 698
SCREEN APPLICATION 701
FILMING AND ALUMINIZING 702
BAKEOUT AND EXHAUST 704
EXHAUST CONSIDERATIONS 704
Pressure Stability
Residual Gas Composition
REFERENCES 707

Color Picture Tube Design and Processing

H. C. Moodey and A. M. Morrell

OPERATIONAL PRINCIPLES OF MAJOR COLOR-TUBE TYPES 709
Accurate Beam Scanning
Beam Switching Near the Phosphor Screen
Direction-Sensitive Screen Assembly
THE GEOMETRY OF SHADOW-MASK TUBE DESIGN 715
Planar Mask-Screen Assembly
Curved Mask-Screen Assembly
ELECTRON-OPTICAL DEVIATIONS FROM THE SIMPLE OPTICAL ANALOGY FOR SHADOW-MASK TUBES 717
Radial Error
Degrouping Errors
Ambient Magnetic Fields
TRIPLE-BEAM GUNS 721
Special Requirements
Description of Gun Types
MECHANICAL DESIGN FEATURES OF SHADOW-MASK TUBES 724
The Shadow-Mask-Frame Assembly
Bulb Design
Gun Design
COMPONENTS FOR SHADOW-MASK TUBES 728
Yoke
Convergence Assembly
Purifying Magnet
Equalizers and Shielding
OPERATIONAL CHARACTERISTICS 729
Drive Characteristics
Magnetic Effects
DESIGN EQUATIONS 732
Definitions
Equations
21CYP22 Parameters
COLOR PICTURE TUBE PROCESSING 733
Panel Assembly
Screening Operation
Funnel
Frit Sealing
Gun Seal
Exhaust
Basing, Aging, and Testing
REFERENCES 734

Circuit Troubles Caused by Unusual Tube Effects

W. E. Babcock

SLEEPING SICKNESS 736
Cathode - Interface Problem
Interface in Entertainment Tubes
BLACKOUT 737
DC SHIFT 739
STRAY EMISSION 739
BULB OR MICA CHARGE 740
SPOOK INTERFERENCE 741
SNIVETS INTERFERENCE 741
REFERENCES 742

Design, Processing, and Application of Industrial-Type Tubes

Some Factors Affecting the Design and Application of Small Power Tubes

J. W. Gaylord

INTRODUCTION 743
- Design of Output Section
- Design of Input Section
- Tube Operation

AMPLIFIER CLASSIFICATION 743
- Current Component Ratios
- Graphical Analysis of Constant-Current Curves
- Fourier Analysis of Constant-Grid-Voltage Curves
- Choice of Operating Point
- Cathode Drive Considerations

CIRCUIT OPERATION 747
- The Q Factor
- Circuit Efficiency
- Determination of E_p and I_p for Lumped Circuits

FREQUENCY EFFECTS 751
- Distributed Circuits
- Simplified Transmission-Line Theory
- Stored Energy in Resonant Transmission Lines
- Determination of E_p and I_p for Distributed Circuits

PHYSICAL DESIGN CONSIDERATIONS 754
- Current Division
- Empirical Data on Cathode-Current Division
- Determination of Cathode Current
- Thermionic Emission
- Amplification of Mu Factor
- Transconductance
- Perveance

DEFINITIONS OF SYMBOLS USED 761

REFERENCES 762

Large-Power-Tube Design, Processing, and Testing

W. P. Bennett

APPLICATIONS AND SERVICE 763

THE CONCEPTUAL DESIGN 763

DETERMINATION OF POWER REQUIREMENTS 764

DETERMINATION OF PLATE VOLTAGE 764

SELECTION OF THE CATHODE 765

COMPUTATION OF THE REQUIRED CATHODE AREA 767

CATHODE CONFIGURATION AND SUPPORT STRUCTURE 767

TYPE AND CONFIGURATION OF TUBE 770
- Frequency Effects in Control Service
- Frequency Effects in RF Generator Service
- Triodes
- Tetrodes
- Electron-Conduction Systems
- Circuit Connections
- Dissipation Considerations

VACUUM ENVELOPES 782
- Electrical Properties of the Ceramic
- Mechanical Properties of Ceramics and Metals
- Bonding Techniques
- Strain Isolation
- Surface Losses
- Assembly Process
- Voltage Standoff
- Materials

MECHANICAL ASPECTS OF TUBE DESIGN 785

EXHAUST AND PROCESSING 786
- Bakeout
- Heating the Electrodes
- High-Voltage Breakdown
- Activation and Aging
- Fault Protection
- Getters and Getter Ion Pumps

TEST AND EVALUATION 788
- Over-all Evaluation
- Diagnostic Tests
- Mechanical Tests
- Circuit Development

REFERENCES 790

Gas Tube Design

H. H. Wittenberg

PROCESSES PRIOR TO CONDUCTION 792
- Collision
- Ionization

BREAKDOWN 795
- Secondary Emission at Cathode
- Paschen's Law

PROCESSES DURING CONDUCTION 797
- Voltage vs. Current
- Voltage vs. Distance
- Sheath
- Plasma

RECTIFIERS 800
- Anode Efficiency
- Cathode Efficiency
- Cathode Design
- Heater Design
- Gas Filling
- Anode Design
- Tube Drop
- Arcback

THYRATRONS 810
- Control Characteristic
- Grid Current Prior to Conduction
- Grid Current During Conduction
- Design

REFERENCES 814

APPENDIX A. BREAKDOWN 816

APPENDIX B. PASCHEN'S LAW 816

APPENDIX C. COLLISION OF ELECTRON AND ION 816

APPENDIX D. BREAKDOWN OF A THYRATRON 817

Design and Processing of Oscillograph Tubes

E. M. Smith

COMPARISON OF MAGNETIC AND ELECTROSTATIC DEFLECTION 818

ELECTRON GUN DESIGN—
ELECTROSTATIC-DEFLECTION TYPES 820
INTERELECTRODE CAPACITANCE 821
DEFLECTING-ELECTRODE SHIELD 821
DEFLECTING-ELECTRODE DESIGN 822
Center of Deflection
Deflection Sensitivity
Electron-Beam Trajectory
POST-DEFLECTION ACCELERATION 823
MAGNETIC DEFLECTION 824
PHOSPHOR SCREEN 824
PROCESSING 824
BIBLIOGRAPHY 825

Phototube Design

J. L. Weaver

PHOTOCATHODES 826
SECONDARY-ELECTRON MULTIPLIERS 827
Multiplier Dynodes
Multiplier Structures
PHOTOTUBE DESIGN TESTS 832
Vacuum and Gas Photodiode Tests
Multiplier Phototube Tests
REFERENCES 836

Camera Tube Design and Processing

B. H. Vine

TARGETS 838
Deposition of Signal Charge on the Target
Scanning, Signal Extraction, and Erasure
High-Velocity Scanning vs. Low-Velocity Scanning
TARGET DESIGN 839
Target Size
Target Surface-Potential Swing
Target Charge
Some Target Relationships
Calculation of Capacitance Lag
Charge Required on an Image Orthicon Target
Noise in the Image Orthicon
Video Amplifier Noise
Capacitance Effects in Wide-Spaced Image Orthicons
ELECTRON OPTICS 842
Focusing
Deflection Aberrations
Beam-Landing Errors
Limitations on Beam-Current Density
MISCELLANEOUS SPURIOUS EFFECTS 844
Sticking and Burn
Vertical Tilt
Flicker
Waterfall Effect
PERFORMANCE CHARACTERISTICS 844
Spectral-Response Characteristic
Transfer Characteristic
Resolution
Noise
Persistence Characteristic
Sensitivity
PROCESSING OF CAMERA TUBES 845
Common Problems
Materials
Mesh Fabrication
Glass Target Fabrication
Photocathodes
REFERENCES 846

Image-Converter Tubes

R. G. Stoudenheimer

HISTORY 848
CONVERSION GAIN 849
SCREEN BACKGROUND 851
Thermionic Emission
Field Emission
ELECTRON LENS SYSTEMS 853
Uniform Electric and Magnetic Fields
Short Magnetic Lens
The Electrostatic Lens
Lens Aberrations
REFERENCES 860

Design and Processing of Storage Tubes

M. D. Harsh

STORAGE TUBES AND INFORMATION PROCESSING 862
DESCRIPTION OF RCA STORAGE TUBES 862
Oscillograph-Type Electrostatic Storage Tube
Radechon
Graphechon
Display Storage Tube
THE STORAGE TARGET 866
The Backplate - Perforated Type
The Backplate - Continuous, Electron-Transparent Type
The Backplate - Continuous, Electron-Opaque Type
Choice of the Dielectric
The Dielectric — Its Preparation
Auxiliary Target-Assembly Elements
Luminescent Screens
ELECTRON-OPTICAL STRUCTURES IN STORAGE TUBES 872
Electron Guns
Collimating Systems
Lens Raster Systems
CONCLUSION 874
REFERENCES 874

Design and Processing of Microwave Magnetrons

F. E. Vaccaro and D. E. Nelson

DESIGN OF THE MAGNETRON INTERACTION SPACE ... 878
Interdependence of Magnetron Components
Modes of the Magnetron
The Equations of Cross-Field Interaction
Interaction-Space Design Equations
Magnetron Scaling
Reduced-Performance Charts; Generalized Hartree Diagrams

MAGNETRON RESONANT CIRCUITS 885
Types of Resonant Circuits
Design of Strapped-Resonator Systems
MAGNETRON TUNING 887
Single Tuning Cavity
Four-Cavity Tunable Magnetrons
MAGNETRON CATHODES 892
OUTPUT CIRCUITS 894
THE MAGNETIC CIRCUIT 894
CONSTRUCTION TECHNIQUES 895
MAGNETRON PROCESSING 896
Pre-exhaust Treatment
Exhaust
Getters
Aging
LIST OF SYMBOLS 896
REFERENCES 897

Traveling-Wave Tube Design

E. E. Bliss

TRAVELING-WAVE TUBE OPERATION

FORWARD-WAVE TRAVELING-WAVE TUBE 898
BACKWARD-WAVE TRAVELING-WAVE TUBE 899
SLOW-WAVE CIRCUITS 900

ANALYSIS OF OPERATION

MATHEMATICAL DESCRIPTION OF A WAVE ON A SLOW-WAVE STRUCTURE 900
MATHEMATICAL DESCRIPTION OF A WAVE ON A SLOW-WAVE STRUCTURE IN THE PRESENCE OF AN ELECTRON BEAM 901
THE DETERMINATION OF Γ 901
CALCULATION OF GAIN 901

FORWARD-WAVE AMPLIFIER DESIGN

THE HELIX AS A SLOW-WAVE STRUCTURE 902
Helix Phase Velocity
Helix Impedance
Limitation of the Helix
OTHER SLOW-WAVE STRUCTURES 907
ELECTRON GUN 909
Convergent-Flow Electron Gun
Parallel-Flow Gun
Hollow-Beam Parallel-Flow Gun
Convergent-Flow Hollow-Beam Gun
Magnetron-Injection Gun
Low-Noise Gun
FOCUSING 912
Magnetic Focusing
Electrostatic Focusing
COUPLERS 918
Coupling to Helices
Coupling to Other Types of Slow-Wave Structures
ATTENUATORS 922
Attenuators for Helices
Attenuators for Other Types of Slow-Wave Structures
COLLECTOR 923
Depressed-Collector Operation
Two-Stage Depressed Collector
Secondary-Emission Suppression
Collector Cooling

APPENDIX

DESIGN PROCEDURE FOR TRAVELING-WAVE TUBES 925
GENERAL TUBE DESIGN 925
SYMBOLS 927
REFERENCES 928